PENDUGAAN RESIKO RELATIF PADA PENDUGAAN AREA KECIL[1]

Kismiantini
Jurusan Pendidikan Matematika
FMIPA Universitas Negeri Yogyakarta

Abstrak

Penduga resiko relatif merupakan statistik yang digunakan untuk mengetahui sebaran suatu penyakit. Penduga resiko relatif sederhana yaitu *standardized mortality ratio* (*SMR*) didasarkan asumsi bahwa banyaknya pengamatan suatu kasus menyebar Poisson. Penduga *SMR* ini tidak lagi dapat diandalkan apabila ukuran contohnya kecil karena akan menghasilkan galat baku yang besar. Metode alternatif untuk menangani masalah tersebut adalah *empirical Bayes* dengan model Poisson-Gamma. Model ini selain menangani kecilnya ukuran contoh juga mampu menangani overdispersi pada model Poisson.

Kata-kata kunci : resiko relatif, penduga *empirical Bayes*

PENDAHULUAN

Penduga resiko relatif merupakan statistik pada pemetaan penyakit yang digunakan untuk mengetahui sebaran penyakit (Clayton & Kaldor 1987). Penduga resiko relatif sederhana adalah *standardized mortality ratio* (*SMR*), yang diperoleh dari asumsi bahwa banyaknya pengamatan suatu kasus menyebar Poisson. Hal tersebut dikarenakan data penyakit berupa data cacahan. Pada pemetaan penyakit, kecilnya ukuran contoh (jumlah kasus berpenyakit) menjadi suatu masalah yang sering dihadapi akibat areanya yang sangat kecil, penyakit jarang terjadi atau keduanya (Law & Haining 2004). Sehingga pendugaan langsung yaitu *standardized mortality ratio* (*SMR*) menjadi tidak dapat diandalkan. Pendugaan area kecil (*small area estimation*) adalah suatu teknik statistika yang mampu menangani permasalahan tersebut, teknik ini berguna untuk menduga parameter subpopulasi (area) yang berukuran contoh kecil (Rao 2003). Salah satu metode alternatifnya adalah metode *empirical Bayes* (*EB*), dengan model yang sering digunakan adalah model Poisson-Gamma (Butar 1991). Selain menangani permasalahan kecilnya ukuran contoh, model ini juga mampu menangani masalah overdispersi pada model Poisson (Gcschlößl & Czado 2006). Pemasukkan peubah penyerta pada model Poisson-Gamma dapat memperbaiki hasil dugaan resiko relatif (Rao 2003). Pada makalah ini akan

[1] Makalah ini disampaikan pada Seminar Nasional Pendidikan Matematika dan Matematika, Jurusan Pendidikan Matematika FMIPA Universitas Negeri Yogyakarta tanggal 24 Nopember 2007

dibahas dua penduga resiko relatif pada pendugaan area kecil menggunakan metode *empirical Bayes* berdasarkan model Poisson-Gamma.

STANDARDIZED MORTALITY RATIO

Standardized mortality ratio (*SMR*) merupakan penduga sederhana resiko relatif pada pemetaan penyakit (Wakefield & Elliott 1999), yang selanjutnya disebut sebagai penduga langsung dalam pendugaan area kecil. *SMR* berguna dalam mengetahui sebaran geografis suatu penyakit. *SMR* ini diperoleh dari asumsi umum pemetaan penyakit bahwa banyaknya pengamatan suatu kasus yaitu $Y_i \overset{ind}{\sim} Poisson(e_i\theta_i)$, dengan e_i menyatakan nilai harapan banyaknya suatu kasus area ke-i dan θ_i adalah resiko relatif area ke-i yang tidak diketahui. Selanjutnya dengan memaksimumkan fungsi peluangnya diperoleh $\hat{\theta}_i = y_i/e_i$, yang merupakan penduga kemungkinan maksimum yang bersifat tak bias.

PENDUGA *EMPIRICAL BAYES* 1

Pendugaan resiko relatif dengan metode *empirical Bayes* pada model poisson-gamma dengan peubah penyerta dikemukakan oleh Wakefield (2006). Pada tahap pertama, diasumsikan bahwa $Y_i \overset{ind}{\sim} Poisson(e_i\mu_i\theta_i)$ dengan $\mu_i = \mu(\underline{x}_i, \underline{\beta})$ menyatakan model regresi sehingga $\underline{x}_i = (x_{1i}, \cdots, x_{pi})^T$ merupakan vektor peubah penyerta tetap, $\underline{\beta} = (\beta_1, \ldots, \beta_p)^T$ merupakan vektor koefisien regresi dan $\theta_i = SMR_i$ adalah penduga langsung *standardized mortality ratio*. Tahap kedua diasumsikan bahwa $\theta_i|\alpha \overset{iid}{\sim} gamma(\alpha,\alpha)$ yang selanjutnya sebagai prior dengan rata-rata 1 dan ragam $1/\alpha$.

Fungsi peluang dari $Y_i \overset{ind}{\sim} Poisson(e_i\mu_i\theta_i)$ adalah

$$f(y_i|\theta_i) = \frac{e^{-e_i\mu_i\theta_i}(e_i\mu_i\theta_i)^{y_i}}{y_i!}, \quad y_i = 0,1,\ldots \tag{1}$$

Fungsi kepadatan peluang dari $\theta_i|\alpha \overset{iid}{\sim} gamma(\alpha,\alpha)$ adalah

$$\pi(\theta_i) = \frac{\alpha^\alpha}{\Gamma(\alpha)} e^{-\alpha\theta_i}\theta_i^{\alpha-1}, \quad \theta_i > 0 \tag{2}$$

Sehingga fungsi bersamanya diperoleh

$$f(y_i, \theta_i) = \frac{e^{-e_i\mu_i\theta_i}(e_i\mu_i\theta_i)^{y_i}}{y_i!}\frac{\alpha^\alpha}{\Gamma(\alpha)} e^{-\alpha\theta_i}\theta_i^{\alpha-1}, \quad y_i = 0,1,\ldots; \theta_i > 0 \tag{3}$$

Selanjutnya fungsi marjinalnya adalah

$$
\begin{aligned}
m(y_i) &= \frac{\alpha^{\alpha}(e_i\mu_i)^{y_i}}{\Gamma(\alpha)y_i!(e_i\mu_i+\alpha)^{y_i+\alpha}}\Gamma(y_i+\alpha) \\
&= \binom{y_i+\alpha-1}{\alpha-1}\left(\frac{\alpha}{e_i\mu_i+\alpha}\right)^{\alpha}\left(1-\frac{\alpha}{e_i\mu_i+\alpha}\right)^{y_i}
\end{aligned}
\tag{4}
$$

Fungsi marjinal diatas merupakan fungsi sebaran binomial negatif dengan rata-rata dan ragam berikut :

$$E\left[Y_i \middle| \underline{\beta},\alpha\right] = e_i\mu_i \tag{5}$$

dan $$Var\left(Y_i \middle| \underline{\beta},\alpha\right) = e_i\mu_i(1+e_i\mu_i/\alpha) \tag{6}$$

Sehingga ragam meningkat sebagai fungsi kuadratik dari rata-rata, dan parameter skala α dapat mengakomodasi overdispersi. Dugaan parameter prior, yaitu $\underline{\hat{\beta}}$ dan $\hat{\alpha}$, dapat diperoleh dari sebaran marjinal $Y_i \middle| \underline{\beta},\alpha \overset{iid}{\sim} \text{binomial negatif}$ menggunakan pendugaan kemungkinan maksimum, yang merupakan penyelesaian dari teknik regresi binomial negatif.

Selanjutnya berdasarkan teorema Bayes maka diperoleh fungsi posterior sebagai berikut :

$$
\begin{aligned}
\pi(\theta_i|y_i,\beta,\alpha) = \frac{f(y_i,\theta_i)}{m(y_i)} &= \frac{\dfrac{e^{-e_i\mu_i\theta_i}(e_i\mu_i\theta_i)^{y_i}}{y_i!}\dfrac{\alpha^{\alpha}}{\Gamma(\alpha)}e^{-\alpha\theta_i}\theta_i^{\alpha-1}}{\dfrac{\alpha^{\alpha}(e_i\mu_i)^{y_i}}{\Gamma(\alpha)y_i!(e_i\mu_i+\alpha)^{y_i+\alpha}}\Gamma(y_i+\alpha)} \\
&= \frac{(e_i\mu_i+\alpha)^{y_i+\alpha}}{\Gamma(y_i+\alpha)}e^{-(e_i\mu_i+\alpha)\theta_i}\theta_i^{\,y_i+\alpha-1},\ \theta_i>0
\end{aligned}
\tag{7}
$$

Sehingga $\theta_i \middle| y_i,\underline{\beta},\alpha \sim gamma(y_i+\alpha, e_i\mu_i+\alpha)$. Dari fungsi posterior tersebut diperoleh penduga Bayes bagi θ_i dan ragam posterior bagi θ_i adalah

$$\hat{\theta}_i^{B}(\underline{\beta},\alpha) = E(\theta_i|y_i,\underline{\beta},\alpha) = (y_i+\alpha)/(e_i\mu_i+\alpha) \tag{8}$$

dan $$V(\theta_i|y_i,\underline{\beta},\alpha) = g_{1i}(\underline{\beta},\alpha,y_i) = (y_i+\alpha)/(e_i\mu_i+\alpha)^2 \tag{9}$$

Penduga *empirical Bayes* bagi θ_i menurut Wakefield (2006) diuraikan menjadi berikut :

$$\hat{\theta}_i^{EB} = \hat{\theta}_i^{B}(\underline{\hat{\beta}},\hat{\alpha}) = \hat{\mu}_i\times(1-w_i)+SMR_i\times w_i \tag{10}$$

dengan $\hat{w}_i = e_i\hat{\mu}_i / (\hat{\alpha} + e_i\hat{\mu}_i)$, $\hat{\mu}_i = \exp(\underline{x}_i^T \underline{\hat{\beta}})$, $\hat{\mu}_i$ adalah nilai harapan resiko relatif ke-*i* yang merupakan penduga tak langsung, $SMR_i = \hat{\theta}_i = y_i / e_i$ adalah penduga langsung (*standardized mortality ratio*) dari θ_i, y_i dan e_i masing-masing menyatakan banyaknya pengamatan dan nilai harapan banyaknya suatu kasus.

PENDUGA *EMPIRICAL BAYES* 2

Wakefield (2006) memberikan alternatif penduga *empirical Bayes* bagi resiko relatif θ_i pada persamaan (10). Langkah pertama dengan mengasumsikan bahwa $Y_i|\theta_i \overset{ind}{\sim} Poisson(e_i\theta_i)$ dan kedua adalah $\theta_i \overset{ind}{\sim} Gamma(\alpha\mu_i, \alpha)$ dengan rata-rata $E(\theta_i) = \mu_i$ dan ragam $V(\theta_i) = \mu_i/\alpha$. Dari model dua tahap ini selanjutnya dapat diperoleh sebaran marjinalnya adalah sebaran binomial negatif dengan rata-rata $E(Y_i|\underline{\beta}, \alpha) = e_i\mu_i$ dan ragam $V(Y_i|\underline{\beta}, \alpha) = e_i\mu_i(1 + e_i/\alpha)$, rata-rata tersebut sama dengan persamaan (5) tetapi ragamnya berbeda. Sehingga untuk menduga $\underline{\beta}$ dan α dapat dilakukan dengan teknik regresi binomial negatif. Berdasarkan teorema Bayes diperoleh fungsi posterior berikut :

$$\begin{aligned}\pi(\theta_i|y_i, \beta, \alpha) = \frac{f(y_i, \theta_i)}{m(y_i)} &= \frac{\dfrac{e^{-e_i\theta_i}(e_i\theta_i)^{y_i}}{y_i!}\dfrac{\alpha^{\alpha\mu_i}}{\Gamma(\alpha\mu_i)}e^{-\alpha\theta_i}\theta_i^{\alpha\mu_i - 1}}{\dfrac{\alpha^{\alpha\mu_i}e_i^{y_i}}{\Gamma(\alpha\mu_i)y_i!(e_i + \alpha)^{y_i + \alpha\mu_i}}\Gamma(y_i + \alpha\mu_i)} \\ &= \frac{(e_i + \alpha)^{y_i + \alpha\mu_i}}{\Gamma(y_i + \alpha\mu_i)}e^{-(e_i + \alpha)\theta_i}\theta_i^{y_i + \alpha\mu_i - 1}, \ \theta_i > 0\end{aligned} \quad (11)$$

Maka $\theta_i|y_i, \underline{\beta}, \alpha \sim gamma(y_i + \alpha\mu_i, e_i + \alpha)$. Dari fungsi posterior tersebut diperoleh penduga Bayes bagi θ_i dan ragam posterior bagi θ_i adalah

$$\hat{\theta}_i^B(\underline{\beta}, \alpha) = E(\theta_i|y_i, \underline{\beta}, \alpha) = (y_i + \alpha\mu_i)/(e_i + \alpha)$$

dan $$V(\theta_i|y_i, \underline{\beta}, \alpha) = g_{1i}(\underline{\beta}, \alpha, y_i) = (y_i + \alpha\mu_i)/(e_i + \alpha)^2 \quad (12)$$

Penduga *empirical Bayes* bagi θ_i menurut Wakefield (2006) diberikan sebagai berikut :

$$\hat{\theta}_i^{EB} = \hat{\theta}_i^B(\underline{\hat{\beta}}, \hat{\alpha}) = \hat{\mu}_i \times (1 - w_i) + SMR_i \times w_i \quad (13)$$

dengan $SMR_i = \dfrac{y_i}{e_i}$, $w_i = e_i/(\hat{\alpha} + e_i)$, dan $\hat{\mu}_i = \exp(\underline{x}^T \underline{\beta})$, $\hat{\mu}_i$ adalah nilai harapan resiko relatif ke-*i* yang merupakan penduga tak langsung, $SMR_i = \hat{\theta}_i = y_i / e_i$ adalah penduga langsung (*standardized mortality ratio*) dari θ_i, y_i dan e_i masing-masing menyatakan banyaknya pengamatan dan nilai harapan banyaknya suatu kasus.

PENERAPAN PADA DATA DINAS KESEHATAN DAN PODES 2005

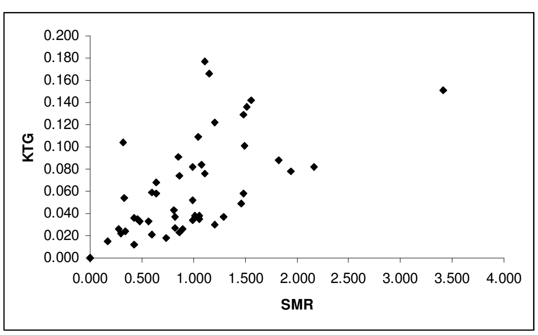
Peubah yang diamati dan menjadi perhatian dalam ilustrasi ini adalah resiko relatif suatu wilayah terjangkit penyakit demam berdarah dengue. Data yang digunakan mengenai data demam berdarah dengue pada 52 kelurahan (area kecil) di kota Bekasi. Data ini berupa banyaknya penderita demam berdarah dengue pada kelurahan ke-*i* (dari data Dinas Kesehatan kota Bekasi tahun 2005), banyaknya penduduk pada kelurahan ke-*i* dan jumlah sekolah pada kelurahan ke-*i* (dari data PODES 2005). Peubah penyerta yang diasumsikan mempengaruhi peubah perhatian adalah jumlah sekolah (SD, SLTP, SMU, SMK negeri dan swasta).

Analisis data menggunakan SAS 9.1 meliputi: PROC GENMOD untuk mendapatkan penduga $\underline{\beta}$ dan α, dan PROC IML untuk mendapatkan penduga *SMR*, *EB* 1 dan *EB* 2 beserta KTG untuk masing-masing penduga. Berikut statistik dari data tersebut :

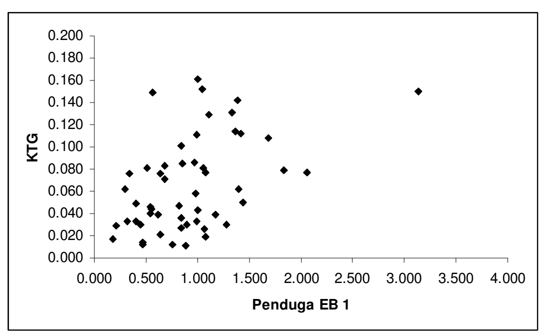
Tabel 1. Pendugaan resiko relatif

Statistik	***SMR***		**Penduga *EB* 1**		**Penduga *EB* 2**	
	RR	**KTG**	**RR**	**KTG**	**RR**	**KTG**
Rata-rata	0.909	0.059	0.926	0.066	0.924	0.052
Simpangan baku	0.621	0.045	0.519	0.042	0.528	0.031

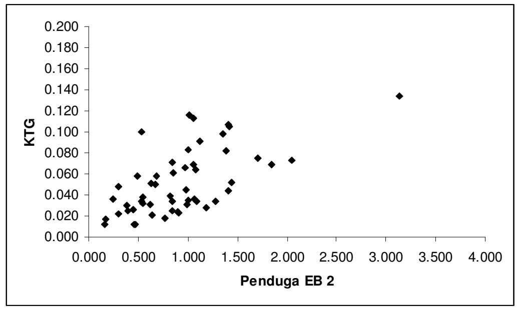
Keterangan : RR : Resiko Relatif; KTG: Kuadrat Tengah



Gambar 1 Plot *SMR* dengan KTG



Gambar 2 Plot Penduga *EB* 1 dengan KTG



Gambar 3 Plot Penduga *EB* 2 dengan KTG

Berdasarkan Tabel 1 dan Gambar 3 dapat diperoleh informasi bahwa penduga resiko relatif dengan metode *EB* 2 memberikan perbaikan pendugaan terhadap hasil penduga *EB* 1 yang ditunjukkan dengan semakin kecilnya nilai kuadrat tengah galat. Penduga *EB* 2 juga menghasilkan ketelitian yang paling baik di antara ketiga penduga, yang diperlihatkan dengan rata-rata nilai kuadrat tengah yang paling kecil. Secara umum dari Gambar 1, 2 dan 3 dapat dilihat bahwa semakin besar nilai penduga resiko relatif maka semakin besar pula nilai kuadrat tengah galatnya.

SIMPULAN

Penduga resiko relatif dengan metode *empirical Bayes* 2 berdasarkan model Poisson-Gamma memberikan ketelitian yang meningkat dibanding penduga dengan metode *empirical Bayes* 1 dan *standardized mortality ratio*.

DAFTAR PUSTAKA

Butar FB. 1991. Empirical Bayes estimation of cancer mortality. Texas: Sam Houston State University.

Clayton D, Kaldor J. 1987. Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics* 43:671-681.

Gschlößl S, Czado C. 2006. Modelling count data with overdispersion and spatial effects. [terhubung berkala]. http://epub.ub.uni-muenchen.de/ [6 Nopember 2007].

Law J, Haining R. 2004. Spatial modeling with discrete response data. University of Cambridge

Marshall RJ. 1991. Mapping disease and mortality rates using empirical Bayes estimators. *Applied Statistics* 40:283-294.

Rao JNK. 2003. *Small Area Estimation.* New York: John Wiley and Sons.

Wakefield J, Elliott P. 1999. Issues in the statistical analysis of small area health data. *Statistics in Medicine* 18:2377-2399.

Wakefield J. 2006. Disease mapping and spatial regression with count data. [terhubung berkala]. http://www.bepress.com/uwbiostat/paper286.pdf [17 Juni 2006].