Forum Statistika dan Komputasi, April 2006, p:12-16
ISSN : 0853-8115

PENERAPAN METODE JACKKNIFE DALAM PENDUGAAN AREA KECIL[1)]

Anang Kurnia dan Khairil A. Notodiputro

Departemen Statistika FMIPA IPB, dan
Center for Statistics and Public Opinion

RINGKASAN

Metode dasar yang sering digunakan dalam menyelesaikan model pendugaan tidak langsung pada SAE adalah BLUP/EBLUP, EB, dan HB. Namun ketidakpuasan sering muncul karena asumsi kelinieran atau sebaran tertentu tidak selalu dipenuhi dalam suatu analisis. Selain itu, penambahan komponen g_2 dan g_3 dari MSE($\hat{\theta}_i^{BP}$) tidak lain adalah upaya untuk mengkoreksi ketidakpastian akibat terlebih dulu melakukan pendugaan terhadap β dan σ^2_υ. Dengan teknik resampling, jackknife berkembang sebagai suatu metode untuk mengkoreksi bias suatu penduga. Penerapan jackknife pada pendugaan area kecil dilakukan untuk mengkoreksi pendugaan MSE. c

Kata Kunci : Pendugaan area *kecil, EBLUP, Jackknife,*

PENDAHULUAN

Pendugaan area kecil (*small area* estimation = SAE) merupakan konsep terpenting dalam pendugaan parameter secara tidak langsung di suatu area yang relatif kecil dalam percontohan survei (*survey sampling*). Metode pendugaan area kecil digunakan untuk menduga karakteristik dari subpopulasi (domain yang lebih kecil). Pendugaan langsung (*direct estimation)* pada sub-populasi relatif tidak memiliki presisi yang memadai karena kecilnya jumlah contoh yang digunakan untuk memperoleh dugaan tersebut.

Alternatif metode lain adalah dengan cara menghubungkan informasi pada area tersebut dengan area lain melalui model yang tepat. Dengan demikian dugaan tersebut merupakan dugaan tidak langsung (*indirect estimation*), dalam arti bahwa dugaan tersebut mencakup data dari domain yang lain. Rao (2003) menyebutkan bahwa prosedur pendugaan. area kecil pada dasarnya memanfaatkan kekuatan area sekitarnya dan sumber data diluar area yang statistiknya ingin diperoleh.

Ketersediaan model pendugaan parameter area kecil akan sangat membantu khususnya bagi BPS dalam menyediakan kebutuhan data dan informasi yang akurat untuk kebutuhan daerah seperti level propinsi, kabupaten/kota atau bahkan kecamatan dengan memanfaatkan keakuratan data pada level nasional. Bagi pemerintah daerah, informasi yang dihasilkan tersebut akan sangat bermanfaat dalam penyusunan sistem perencanaan, pemantauan dan penilaian pembangunan daerah atau kebijakan penting lainnya tanpa harus **mengeluarkan** biaya **besar** untuk **mengumpulkan** data sendiri.

MODEL PENDUGAAN AREA KECIL

Model dasar pada pendugaan area kecil yang dikembangkan dan dapat dipelajari melalui beberapa literatur dapat dibedakan menjadi basic *area level* dan *basic* unit *level model*. *Basic area* didasarkan pada ketersediaan data pendukung yang hanya ada untuk level area tertentu, katakan $\mathbf{x}_i = (x_{1i}, x_{2i}, \ldots, x_{pi})'$ yang akan digunakan untuk membangun model $\theta_i = \mathbf{x}_i'\beta + \upsilon_i$ dengan $\upsilon_i \sim N(0, \sigma^2_\upsilon)$. Suatu model yang menggabungkan model berdasarkan penarikan contoh yang bersesuaian $\hat{\theta}_i = \theta_i + e$, dimana $\hat{\theta}_i$ adalah penduga langsung bagi θ_i dan $e_i|\theta_i \sim N(0, \sigma^2_{ei})$ serta σ^2_{ei} yang diketahui dengan model $\theta_i = \mathbf{x}_i'\beta + \upsilon_i$ untuk menghasilkan model gabungan $\hat{\theta}_i = \mathbf{x}_i'\beta + \upsilon_i + e_i$ Penjelasan detail dapat dilihat pada Saei dan Chambers(2003); Rao(2003); Rao(1999).

Basic unit, memerlukan data pendukung yang bersesuaian secara individu dengan data respon, misal $\mathbf{x}_{ij} = (x_{1ij}, x_{2ij}, \ldots, x_{pij})'$, sehingga bisa dibangun suatu model regresi tersarang $y_{ij} = \mathbf{x}_{ij}^T\beta + \upsilon_i + e_{ij}$ dengan $\upsilon_i \sim N(0, \sigma^2_\upsilon)$ dan $e_{ij} \sim N(0, \sigma^2_e)$. Namun demikian, model yang diperoleh biasanya merupakan model yang komplek sehingga memerlukan teknik yang relatif lebih rumit dalam penyelesaiannya. Penjelasan detail dapat dilihat pada Rao (2003).

Fay dan Herriot (1979) mengembangkan model $y_i = x_i'\beta + \upsilon_i + e_i$ sebagai dasar dalam pengembangan pendugaan area kecil. Untuk selanjutnya diasumsikan bahwa β dan σ^2_υ tidak diketahui, tetapi σ^2_{ei} (i = 1, 2, ..., m) diketahui. Penduga terbaik (BP) bagi $\theta_i = x_i'\beta + \upsilon_i$ jika β dan σ^2_υ diketahui adalah

$$\hat{\theta}_i^{BP} = \hat{\theta}_i(y_i \mid \beta, \sigma^2_\upsilon) = x_i'\beta + (1 - B_i)(y_i - x_i'\beta) \quad (1)$$

dengan $B_i = \sigma^2_{ei} / (\sigma^2_\upsilon + \sigma^2_{ei})$ untuk i = 1, 2, ..., m dan $MSE(\hat{\theta}_i^{BP}) = Var(\theta_i \mid y_i, \beta, \sigma^2_\upsilon) = (1 - B_i)\sigma^2_{ei} = g_{1i}(\sigma^2_\upsilon)$.

Dalam praktek, baik β maupun σ^2_υ biasanya tidak diketahui sehingga untuk kasus σ^2_υ diketahui, β dapat diduga dengan metode kemungkinan maksimum atau metode momen $\beta^* = \hat{\beta}_i(\sigma^2_\upsilon) = (X'V^{-1}X)^{-1}X'V^{-1}Y$ dengan $V = Diag(\sigma^2_\upsilon + \sigma^2_{e1}, \sigma^2_\upsilon + \sigma^2_{e2}, ..., \sigma^2_\upsilon + \sigma^2_{em})$. Kemudian dengan mensubstitusi β oleh β^* pada $\hat{\theta}_i^{BP}$, maka diperoleh

$$\hat{\theta}_i^{BLUP} = \hat{\theta}_i(y_i \mid \sigma^2_\upsilon) = x_i'\beta^* + (1 - B_i)(y_i - x_i'\beta^*) \quad (2)$$

Menurut Ghosh dan Rao (1994) $MSE(\hat{\theta}_i^{BLUP}) = g_{1i}(\sigma^2_\upsilon) + g_{2i}(\sigma^2_\upsilon)$, dengan $g_{2i}(\sigma^2_\upsilon) = (\sigma^2_{ei})^2/(\sigma^2_\upsilon + \sigma^2_{ei})\,[x_i'(X'V^{-1}X)^{-1}x_i]$. Jika terlebih dahulu σ^2_υ diduga oleh s^2_υ baik menggunakan metode ML, REML ataupun momen sehingga dengan mensubstitusi β oleh $\hat{\beta}$ dan σ^2_υ oleh s^2_υ terhadap penduga BLUP ($\hat{\theta}_i^{BLUP}$), maka akan diperoleh suatu bentuk penduga baru

$$\hat{\theta}_i^{EBLUP} = \hat{\theta}_i(y_i \mid s^2_\upsilon) = x_i'\hat{\beta} + (1 - \hat{B}_i)(y_i - x_i'\hat{\beta}) \quad (3)$$

Jika didefinisikan MSE dari $\hat{\theta}_i^{EBLUP}$ adalah

$$MSE(\hat{\theta}_i^{EBLUP}) = E(\hat{\theta}_i^{EBLUP} - \theta_i)^2$$
$$= Var(\hat{\theta}_i^{EBLUP}) + (Bias\ \hat{\theta}_i^{EBLUP})^2$$

dan berdasarkan Kacker dan Harville (1984), persamaan tersebut dapat diuraikan menjadi

$$MSE(\hat{\theta}_i^{EBLUP}) = MSE(\hat{\theta}_i^{BLUP}) + E(\hat{\theta}_i^{EBLUP} - \hat{\theta}_i^{BLUP})^2$$
$$= H_{1i}(\sigma^2_\upsilon) + H_{2i}(\sigma^2_\upsilon) \quad (4)$$

dengan

$$H_{1i}(\sigma^2_\upsilon) = MSE(\hat{\theta}_i^{BLUP}) = g_{1i}(\sigma^2_\upsilon) + g_{2i}(\sigma^2_\upsilon)$$
$$H_{2i}(\sigma^2_\upsilon) = E(\hat{\theta}_i^{EBLUP} - \hat{\theta}_i^{BLUP})^2$$

Prasad dan Rao (1990) menggunakan ekspansi deret Taylor untuk menduga $MSE(\hat{\theta}_i^{EBLUP})$ dan diperoleh $MSE(\hat{\theta}_i)^{PR} = g_{1i}(s^2_\upsilon) + g_{2i}(s^2_\upsilon) + 2g_{3i}(s^2_\upsilon)$ dengan

$$g_{3i}(s^2_\upsilon) = \frac{2\sigma^2_{ei}}{m^2(s^2_\upsilon + \sigma^2_{ei})^2}\sum_{j=1}^{m}(s^2_\upsilon + \sigma^2_{ej})^2 .$$

Ketidakpuasan kemudian muncul karena asumsi kelinieran tidak selalu dipenuhi dalam suatu analisis. Selain itu, penambahan komponen g_2 dan g_3 dari $MSE(\hat{\theta}_i^{BP})$ menjadi $MSE(\hat{\theta}_i^{BLUP})$ dan $MSE(\hat{\theta}_i^{EBLUP})$ tidak lain adalah upaya untuk mengoreksi ketidakpastian akibat terlebih dulu melakukan pendugaan terhadap β dan σ^2_υ.

PENDEKATAN JACKKNIFE

Metode jackknife pertama diperkenalkan oleh Tukey (1958). Jackknife kemudian berkembang sebagai suatu metode untuk mengkoreksi bias suatu penduga. Dengan melakukan penghapusan terhadap observasi ke-i untuk i = 1, 2, ..., m dan kemudian dilakukan pendugaan parameter, misal $\hat{\theta}_{(i)}$, maka penduga bias diduga oleh

$$bias(\hat{\theta}) = (m-1)[\,\hat{\theta}_{(.)} - \hat{\theta}\,] \quad (5)$$

dengan $\hat{\theta}_{(.)} = m^{-1}\,\Sigma_i\hat{\theta}_{(i)}$. Penduga jackknife diperoleh : $\hat{\theta}_{jack} = \hat{\theta} - bias(\hat{\theta})$ dan $v_{jack}(\hat{\theta}) = (n-1)/n\,\Sigma_i[\hat{\theta}_{(i)} - \hat{\theta}]^2$.

Penerapan jackknife pada pendugaan area kecil dilakukan untuk mengkoreksi pendugaan MSE mengingat formula MSE seperti yang telah diuraikan dalam sub bab sebelumnya. Persamaan (4) setara dengan $g_{1i}(s^2_\upsilon)$ + bias jika σ^2_υ diduga. Bias bagi $g_{1i}(s^2_\upsilon)$ bersifat $O(m^{-1})$ (Prasad dan Rao (1990); Lahiri dan Rao (1995)). Dengan menggunakan jackknife $MSE(\hat{\theta}_i^{BLUP})$ diduga oleh

$$b_i^{jack} = g_{1i}(s^2_\upsilon) - (m-1)/m\,\Sigma_u[g_1(s^2_{\upsilon(-u)}) - g_1(s^2_\upsilon)] \quad (6)$$

dengan

$$s^2_\upsilon = (m-1)^{-1}\,\Sigma_i(y_i - y_{bar})^2 - \sigma^2_e$$
$$y_{bar} = m^{-1}\,\Sigma_i\,y_i.$$
$$s^2_{\upsilon(-u)} = (m-2)^{-1}\,\Sigma_{i\neq u}(y_i - y_{bar(-u)})^2 - \sigma^2_e$$
$$y_{bar(-u)} = (m-1)^{-1}\,\Sigma_{i\neq u}\,y_i$$

Adapun $E(\hat{\theta}_i^{EBLUP} - \hat{\theta}_i^{BLUP})^2$ diduga oleh

$$E_i^{jack} = (m-1)/m\,\Sigma_u[\hat{\theta}_i^{EBLUP}{}_{(-u)} - \hat{\theta}_i^{BLUP}]^2 \quad (7)$$

dengan $\hat{\theta}_i{}^{EBLUP}{}_{(-u)} = x_i' \hat{\beta}_{(-u)} + (1 - \hat{B}_{i\,(-u)})(y_i - x_i' \hat{\beta}_{(-u)})$. Dengan demikian, penduga MSE($\hat{\theta}_i{}^{EBLUP}$) dapat didekati ole $mse_i^{jack} = b^{jack} + E_i^{jack}$.

PENERAPAN PADA DATA BPS

Peubah yang diamati dan menjadi perhatian dalam ilustrasi ini adalah rata-rata pengeluaran perkapita rumah tangga. Sumber data yang digunakan adalah SUSENAS 2003 dengan materi informasi berbasis rumah tangga, serta PODES 2003 sebagai sumber data peubah penyerta.
Peubah penyertanya adalah peubah-peubah yang diasumsikan mempengaruhi dan atau menggambarkan pengeluaran rumah tangga pada suatu wilayah, meliputi: persentase rumah tangga prasejahtera dan sejahtera 1, persentase pengangguran, persentase rumah tangga pelanggan listrik PLN, persentase rumah tangga pelanggan telpon, dan persentase rumah tangga tani.

Analisis menggunakan SAS 9.1.3 meliputi: proc mixed untuk pendugaan EBLUP dengan REML, proc tabulate untuk pendugaan langsung dan pemrograman menggunakan proc IML untuk pendugaan dengan metode jackknife. Hasil analisis disajikan pada tabel berikut.

Kajian empirik memperlihatkan bahwa ada potensi yang cukup baik untuk mengembangkan penerapan konsep jackknife dalam mengkoreksi bias pendugaan MSE pada kasus pendugaan area kecil. Keunggulan yang diberikan oleh jackknife pada kajian empirik ini adalah bebas asumsi sebaran karena analisis menggunakan metode momen, walaupun Jiang (1996) memperlihatkan bahwa pendugaan komponen ragam, σ^2_u, pada model linear campuran dengan menggunakan REML akan menghasilkan penduga yang konsisten walaupun asumsi kenormalan tidak sepenuhnya dipenuhi.

Sejalan juga dengan yang telah diperlihatkan Kurnia dan Notodiputro (2005a dan 2005b), pendugaan berdasarkan *design-based* untuk kasus data Susenas di Jawa Barat relatif memberikan hasil yang baik, walaupun tidak selalu yang terbaik. Kasus ini memberikan dugaan awal

Tabel 1. Pendugaan pengeluaran per kapita (x Rp.100.000,-) berdasarkan design-based, model-based : REML dan Jackknife

Kabupaten/Kota	Design		EBLUP-REML		EBLUP-Jackknife	
	Theta_hat	MSE	Theta_hat	MSE	Theta_hat	MSE
Kab. Bogor	2.3785	0.0397	2.2656	0.1798	2.1965	0.0301
Kab. Sukabumi	1.8006	0.0277	1.8650	0.1700	1.8184	0.0328
Kab. Cianjur	1.6890	0.0307	1.9152	0.2161	1.9043	0.0388
Kab. Bandung	2.1101	0.0310	2.0881	0.3769	2.0918	0.0397
Kab. Garut	1.6391	0.0277	1.8882	0.1525	1.8456	0.0351
Kab. Tasikmalaya	1.8493	0.0417	1.8667	0.2992	1.8539	0.0426
Kab. Ciamis	1.8226	0.0318	1.9729	0.1911	1.9523	0.0308
Kab. Kuningan	1.8575	0.0307	1.9364	0.2693	1.9051	0.0321
Kab. Cirebon	1.7127	0.0259	1.8571	0.3597	1.7925	0.0378
Kab. Majalengka	1.9127	0.0348	1.8875	0.0934	1.8496	0.0287
Kab. Sumedang	2.3590	0.0517	2.0505	0.3405	2.0378	0.0446
Kab. Indramayu	2.3934	0.0968	1.9744	0.9097	1.7752	0.1138
Kab. Subang	2.1028	0.0380	2.0264	0.5109	1.9994	0.0464
Kab. Purwakarta	2.2545	0.0351	2.1985	0.2746	2.1630	0.0305
Kab. Karawang	2.1562	0.0361	2.1833	0.1350	2.1277	0.0374
Kab. Bekasi	2.6205	0.0461	2.5072	0.1291	2.4513	0.0384
Kota Bogor	3.4028	0.1011	3.4142	0.7821	3.3213	0.0546
Kota Sukabumi	2.4402	0.0550	2.7519	0.5052	2.7760	0.0788
Kota Bandung	3.4760	0.0776	3.4625	0.5100	3.4401	0.0446
Kota Cirebon	3.0606	0.0967	3.0706	0.8145	3.0991	0.0573
Kota Bekasi	3.7935	0.1062	4.0274	1.4244	4.0369	0.0792
Kota Depok	4.0318	0.1415	3.6538	0.8687	3.6225	0.0719

bahwa ukuran contoh untuk area kabupaten/kota di Jawa Barat cukup memadai untuk digunakan dalam pendugaan langsung.

Temuan-temuan empirik sampai dengan saat ini menuntun untuk dilakukan kajian lebih lanjut baik dari sisi design-based misal dengan mengkaji lebih lanjut kecukupan ukuran contoh dan evaluasi pembobot maupun dari sisi model-based dengan mengembangkan dan atau memperbaiki metode-metode yang sudah ada.

Tabel 2. Ukuran Contoh Susenas 2003 di Propinsi Jawa Barat

Kabupaten/Kota	Ukuran Contoh
Kab. Bogor	1167
Kab. Sukabumi	908
Kab. Cianjur	924
Kab. Bandung	1347
Kab. Garut	896
Kab. Tasikmalaya	928
Kab. Ciamis	864
Kab. Kuningan	639
Kab. Cirebon	864
Kab. Majalengka	702
Kab. Sumedang	654
Kab. Indramayu	832
Kab. Subang	768
Kab. Purwakarta	734
Kab. Karawang	817
Kab. Bekasi	799
Kota Bogor	607
Kota Sukabumi	479
Kota Bandung	942
Kota Cirebon	480
Kota Bekasi	821
Kota Depok	656

PUSTAKA

Fay, R.E. and Herriot, R.A., 1979, "Estimates of income for small places: an application of James-Stein procedures to Census data". *Journal of the American Statistical Association*, Vol. 74, p.269-277.

Ghosh, M. dan J.N.K. Rao. 1994. Small Area Estimation: An Appraisal. Statistical Science, Vol.9 No.1, p: 55 – 98.

Jiang, J. 1996. REML estimation : asymptotic behavior and related topics. Annals of Statistics, 24, p: 255 – 286.

Kurnia, A. dan K.A. Notodiputro, 2005a, General Linear Mixed Model pada *Small Area Estimation*, Makalah disampaikan pada Seminar Nasional Matematika, UI Depok, 30 Juli 2005.

Kurnia, A. dan K.A. Notodiputro, 2005b, Aplikasi Metode Bayes Pada *Small Area Estimation*, Makalah disampaikan pada Seminar Nasional Statistika VII, ITS Surabaya, 26 Nopember 2005.

Lahiri, P.A. dan Rao, J.N.K. 1995. Robust estimation of mean squared error of small area estimator. *Journal of the American Statistical Association*, 82, p: 758 – 766.

Prasad, N.G.N. dan Rao, J.N.K., 1990, The estimation of mean squared errors of small area estimatimators. Journal of the American Statistical Association, 85, 163-171.

Rao, J.N.K., 2003, *Small Area Estimation*, New York : John Wiley and Sons.

Rao, J.N.K., 1999, Some Recent Advances in Model-Based Small Area Estimation, Survey Methodology, Vol.25 No.2, 175-186.

Saei, A. dan R. Chambers, 2003, "Small Area Estimation: A Review of Methods Based on the Application of Mixed Models", S³RI Methodologi Working Paper M03/16, University of Southampton, UK.

⊗