

Evaluasi Model Pendugaan Area Kecil dengan Penambahan Informasi Gerombol

RAY SASTRI¹, ARBI SETIYAWAN²

^{1,2}Politeknik Statistika Sekolah Tinggi Ilmu Statistik, ²Pusdiklat Badan Pusat Statistik
Email: raysastri@stis.ac.id

ABSTRAK

Di setiap provinsi, wilayah kecil seperti desa selalu punya peran dan kontribusi yang besar dalam proses pembangunan dan pertumbuhan ekonomi. Untuk mendukung program tersebut, pemerintah membutuhkan data yang akurat pada level desa agar segala sumber daya dapat dimanfaatkan seefisien mungkin. Namun untuk desa yang memiliki jumlah sampel kecil atau nirsampel, pendugaan langsung tidak sesuai digunakan karena akan menghasilkan galat baku yang sangat besar (Rao, 2013). Alternatifnya adalah melakukan pendugaan tidak langsung melalui model pendugaan area kecil. Dalam model pendugaan area kecil, kita perlu menambahkan pengaruh area ke dalam model. Model yang dibangun dalam penelitian ini adalah model EBLUP dasar dan model EBLUP modifikasi dengan menambahkan informasi gerombol. Berdasarkan nilai perubahan MSE dan relatif bias dalam kajian simulasi, dapat disimpulkan bahwa area nirsampel terbanyak yang bisa ditangani oleh model EBLUP dengan penambahan informasi gerombol adalah 20%.

Kata Kunci: Model pendugaan area kecil, nirsampel, model EBLUP, gerombol.

1. PENDAHULUAN

Di setiap provinsi, wilayah kecil seperti desa selalu punya peran dan kontribusi yang besar dalam proses pembangunan dan pertumbuhan ekonomi. Sejak terbitnya undang-undang tentang desa pada tahun 2014, pemerintah daerah berusaha semaksimal mungkin membuat program-program berbasis desa terutama yang berhubungan dengan peningkatan kesejahteraan masyarakat. Untuk mendukung program tersebut, pemerintah membutuhkan data yang akurat pada level desa agar segala sumber daya dapat dimanfaatkan seefisien mungkin.

Beberapa survei yang dilakukan oleh Badan Pusat Statistik (BPS) mempunyai level estimasi provinsi atau kabupaten. Jumlah sampel tidak mencukupi untuk melakukan pendugaan sampai ke tingkat desa. Beberapa desa bahkan tidak terpilih sebagai sampel atau nirsampel. Untuk desa yang memiliki jumlah sampel kecil atau nirsampel, pendugaan langsung tidak sesuai digunakan karena akan menghasilkan galat baku yang sangat besar (Rao, 2013). Alternatifnya adalah melakukan pendugaan tidak langsung melalui model pendugaan area kecil.

Dalam model pendugaan area kecil, kita perlu menambahkan pengaruh area ke dalam model. Tetapi informasi ini tidak tersedia jika areanya nirsampel. Jika pengaruh acak ini diabaikan, maka sangat mungkin terjadi bias yang besar. Anisa (2014) mencoba menambahkan informasi gerombol sebagai pengaruh acak area dan menyimpulkan bahwa model dengan penambahan informasi gerombol memiliki kemampuan prediksi yang lebih baik dibandingkan dengan model baku untuk area nirsampel. Dalam penelitiannya, Anisa melakukan simulasi dengan menggunakan wilayah nirsampel sebanyak 25 persen. Hasil dari penelitian ini kemudian digunakan untuk menyelesaikan kasus dengan jumlah area nirsampel sebanyak sepuluh persen. Sementara itu, desa yang nirsampel pada survei-survei BPS bisa lebih dari 25 persen. Pada Survei Kemiskinan di Provinsi Kepulauan Bangka Belitung yang dilakukan oleh mahasiswa STIS, desa nirsampel bahkan mencapai 65 persen. Untuk kasus area nirsampel yang besar, belum diketahui apakah metode penambahan informasi gerombol yang diajukan Anisa (2014) ini masih baik untuk digunakan atau tidak. Sehingga perlu dilakukan penelitian sampai sejauh mana metode ini mampu memberikan dugaan yang baik. Oleh karena itu, pertanyaan yang ingin dijawab dalam penelitian ini adalah berapa besar area nirsampel yang bisa ditangani oleh model dengan penambahan informasi gerombol.

2. METODOLOGI PENELITIAN

Pendugaan Area Kecil

Model pendugaan area kecil dapat dibedakan menjadi dua jenis: (i) Model level area, digunakan ketika informasi tambahan tersedia hanya pada level area. Model ini menggunakan peubah respon yang didapatkan dari pendugaan langsung di wilayah kecil dan peubah penjelas berbasis area. Model ini kemudian disebut sebagai model Fay-Herriot. (ii) Model level unit, digunakan ketika tersedia informasi pada level unit baik peubah respon maupun peubah penjelas. Model ini kemudian dikenal sebagai model Battese, Harter dan Fuller. Jika data pada level unit tidak tersedia, pendugaan area kecil dilakukan dengan model level area.

Model Empirical Best Linear Unbiased Predictor (EBLUP) untuk Y_i adalah sebagai berikut:

$$\hat{Y}_i^{EBLUP} = \mathbf{x}_i^T \hat{\boldsymbol{\beta}} + \hat{u}_i = \hat{\gamma}_i y_i + (1 - \hat{\gamma}_i) \mathbf{x}_i^T \hat{\boldsymbol{\beta}} \quad (1)$$

Dimana

$$\hat{\boldsymbol{\beta}} = \left(\sum_i (\sigma_{ei}^2 + \hat{\sigma}_u^2)^{-1} \mathbf{x}_i \mathbf{x}_i^T \right)^{-1} \left(\sum_i (\sigma_{ei}^2 + \hat{\sigma}_u^2)^{-1} \mathbf{x}_i y_i \right)$$

$$\hat{u}_i = \hat{\gamma}_i (y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}})$$

$$\hat{\gamma}_i = \frac{\hat{\sigma}_u^2}{\sigma_{ei}^2 + \hat{\sigma}_u^2}$$

Dengan varians dari model EBLUP adalah sebagai berikut

$$MSE \left[\hat{Y}_i^{EBLUP} (\hat{\boldsymbol{\beta}}, \hat{\sigma}_u^2) \right] = E \left[\hat{Y}_i^{EBLUP} (\hat{\boldsymbol{\beta}}, \hat{\sigma}_u^2) - Y_i \right]^2 = g_{1i} + g_{2i} + g_{3i}$$

Dimana $g_{2i} = (1 - \gamma_i)^2 \mathbf{x}_i^T \text{var}(\hat{\boldsymbol{\beta}}) \mathbf{x}_i$ dengan $\text{var}(\hat{\boldsymbol{\beta}}) = \left(\sum_i (\sigma_{ei}^2 + \sigma_u^2)^{-1} \mathbf{x}_i \mathbf{x}_i^T \right)^{-1}$ adalah

pengembangan MSE karena menduga $\boldsymbol{\beta}$. Dan $g_{3i} = \frac{\sigma_{ei}^4}{(\sigma_{ei}^2 + \sigma_u^2)^3} \text{Var}(\hat{\sigma}_u^2)$ adalah

pengembangan MSE karena menduga σ_u^2 .

Demikian sehingga penduga dari MSE adalah :

$$\widehat{MSE} \left[\hat{Y}_i^{EBLUP} (\hat{\boldsymbol{\beta}}, \hat{\sigma}_u^2) \right] = g_{1i}(\hat{\sigma}_u^2) + g_{2i}(\hat{\sigma}_u^2) + g_{3i}(\hat{\sigma}_u^2)$$

Penduga EBLUP pada area nirsampel ke- i^* adalah sebagai berikut:

$$\hat{Y}_{i^*}^{EBLUP} = \mathbf{x}_{i^*}^T \hat{\boldsymbol{\beta}} \quad (2)$$

Penduga EBLUP pada area nirsampel ke- i^* ditunjukkan pada model (1). Pengaruh acak area hanya dimiliki oleh area sampel, sedangkan area nirsampel tidak diketahui pengaruh acak areanya. Hal ini karena sampling error hanya ada untuk area sampel. Sedangkan area nirsampel tidak memiliki sampling error. Dalam pendugaan pada area nirsampel, karena tidak ada informasi pengaruh acak area, maka pengaruh ini sering diabaikan. Sehingga modelnya menjadi seperti berikut

$$\hat{Y}_{i^*}^{EBLUP} = \mathbf{x}_{i^*}^T \hat{\boldsymbol{\beta}} \quad (3)$$

Annisa (2014) melakukan modifikasi model (3) untuk level unit dengan memasukkan informasi gerombol untuk menduga area nirsampel. Area sampel dan area nirsampel secara bersama-sama digerombolkan berdasarkan peubah X nya sehingga diketahui gerombol untuk setiap wilayah baik wilayah sampel maupun nirsampel. Selanjutnya pada area sampel, pengaruh acak area yang telah diketahui kemudian dirata-ratakan per gerombol. Informasi rata-rata pengaruh acak area per gerombol ini kemudian dimasukkan ke dalam model sebagai penduga dari pengaruh acak area. Penambahan tersebut dinyatakan sebagai

$$\bar{u}_{(k)} = \frac{1}{m_k} \sum_{i=1}^{m_k} \hat{u}_i$$

Dengan m_k adalah banyaknya area sampel pada gerombol ke- k . Sehingga model pendugaan untuk area nirsampel adalah sebagai berikut:

$$\hat{Y}_{i^*,j^*,k}^{EBLUP} = \mathbf{x}_{i^*,j^*,k}^T \hat{\boldsymbol{\beta}} + \tilde{u}_{(k)}$$

Dalam penelitian ini, model yang dibangun adalah model level area. Meskipun model yang dibangun berbeda levelnya dengan model Anisa, model modifikasi ini masih relevan digunakan dalam penelitian ini. Secara sederhana, model EBUP modifikasi dituliskan sebagai berikut:

$$\hat{Y}_{i^*,k} = \mathbf{x}_{i^*}^T \hat{\boldsymbol{\beta}} + \tilde{u}_{(k)}$$

Dimana i^* adalah area nirsampel dan $\tilde{u}_{(k)}$ adalah rata-rata pengaruh acak area pada gerombol ke- k .

Kajian Simulasi

Model dasar yang digunakan dalam penelitian ini adalah model EBLUP dasar dan model EBLUP modifikasi dengan penambahan informasi gerombol seperti yang dikembangkan oleh Anisa (2014). Proses simulasi dilakukan untuk mengukur berapa banyak area nirsampel yang mampu ditangani oleh model modifikasi tersebut. Proses simulasi dilakukan dengan langkah-langkah berikut ini:

- 1) Membangun populasi yang terdiri dari 1000 area. Populasi ini tidak memperhatikan pola penggerombolan. Simulasi ini menggunakan satu peubah respon Y dan satu peubah penyerta X . Peubah respon yang diamati adalah kombinasi linier dari peubah penyerta X , pengaruh acak area dan galat sampel. Parameter β_0 ditentukan sebesar 1 dan β_1 sebesar 0.5. Peubah penyerta X dibangkitkan dari sebaran uniform (0,100), sedangkan komponen acak e dibangkitkan dari sebaran normal(0,0.15) dan v dibangkitkan dari sebaran normal(0,0.5).
- 2) Mengambil sampel acak dari populasi yang sudah dibangkitkan pada langkah (1) yaitu berturut-turut sebanyak 10 persen, 20 persen, 30 persen, 40 persen, 50 persen, 60 persen.
- 3) Melakukan pemodelan dan menduga nilai tengah area nirsampel. Proses ini dilakukan menggunakan model dasar EBLUP dan model EBLUP modifikasi.
- 4) Mengulangi proses pada langkah (1) sampai langkah (4) sebanyak 1000 kali. sehingga dapat dihitung relative bias dan MSE.
- 5) Mengevaluasi model berdasarkan relatif bias dan MSE.

3. HASIL DAN PEMBAHASAN

Data populasi dibangkitkan untuk peubah respon dan seluruh peubah penyerta beserta sampling error dan pengaruh acak area. Data populasi tersebut kemudian dibagi menjadi dua, yaitu data sampel dan data non sampel. Data sampel digunakan untuk membangun model prediksi. Data non sampel didapatkan dengan cara menghilangkan sebagian data populasi untuk mendapatkan daerah non sampel 10 % - 60 %. Model yang dibangun dalam penelitian ini adalah model EBLUP dasar dan model EBLUP modifikasi dengan menambahkan informasi gerombol. Dari hasil simulasi didapatkan hasil pada Tabel 1 sebagai berikut.

Model dengan jumlah area nirsampel sebesar 10 persen memiliki MSE yang paling kecil yaitu 0.539 dan MSE tertinggi pada wilayah nirsampel 60%. Ini hal yang wajar karena semakin banyak wilayah nirsampel, semakin tinggi MSE-nya. Jika kita lihat perbedaan MSE, wilayah nirsampel 20% dengan 10 % sebesar 8.06 sedangkan perbedaan wilayah nirsampel 30% dengan 20% menurun seanyak 4.81. Perbedaan MSE pada wilayah nirsampel 40% sangat kecil bedanya dengan wilayah 50% dan 60%. Sehingga dapat dikatakan bahwa 40% adalah jumlah wilayah nirsampel yang paling banyak yang bisa ditangani oleh model EBLUP modifikasi. Dari nilai relatif bias pada Tabel 2, nilai yang paling mendekati nol adalah pada wilayah nirsampel 20% yaitu sebesar 0,02768. Dengan kata lain, wilayah nirsampel terbanyak yang bisa ditangani oleh model adalah 20%. Berdasarkan nilai perubahan MSE dan relatif bias dapat disimpulkan bahwa area nirsampel terbanyak yang bisa ditangani oleh model EBLUP dengan penambahan informasi gerombol adalah 20%.

Tabel 1. MSE dan Relatif Bias Pada Area Non Sampel dengan Model EBLUP Modifikasi

Jumlah Area Nirsampel	MSE	Δ MSE (%)	Relatif Bias
10%	0.53980		0.29800
20%	0.58333	8.06348	0.02768
30%	0.61141	4.81443	-0.40626
40%	0.62808	2.72593	-0.28396
50%	0.63809	1.59374	0.22949
60%	0.63800	-0.01429	0.26091

4. KESIMPULAN DAN SARAN

Berdasarkan nilai perubahan MSE dan relatif bias dalam kajian simulasi, dapat disimpulkan bahwa area nirsampel terbanyak yang bisa ditangani oleh model EBLUP dengan penambahan informasi gerombol adalah 20%.

5. DISKUSI

Dalam kajian simulasi di penelitian ini, data dibangkitkan tanpa memperhatikan pola penggerombolan. Padahal kenyataannya, gerombol bisa terpisah sempurna atau juga tidak terpisah sempurna. Selain itu, model EBLUP modifikasi dengan menggunakan informasi gerombol yang digunakan hanya satu dari lima model EBLUP modifikasi yang diajukan oleh Anisa (2014). Hal ini dapat menjadi masukan bagi penelitian yang akan datang.

DAFTAR PUSTAKA

- Anisa R, Kurnia A, Indahwati. 2014. Cluster Information of Non-sampled Area in Small Area Estimation. *IOSR Journal of Mathematics* 10(1):15-19.
- Hamdani, Rifki. 2015. Pendugaan Area Kecil Angka Melek Huruf Pada Tingkat Kecamatan di Kabupaten Donggala Dengan Metode Bayes Berhirarki. Bogor: Institut Pertanian Bogor [Tesis]
- Istiqomah, Nurul. 2015. *Penggerombolan Kabupaten/Kota Berdasarkan Faktor Stunting Menggunakan Metode Penggerombolan Dua Langkah Untuk Data Campuran*. Bogor : Institut Pertanian Bogor [Tesis]
- Lahiri, P. and Rao, J.N.K.. 1995. Robust Estimation of Mean Square Error of Small Area Estimation. *Journal of American Statistical Association*. 90: 758-766.
- Mattjik AA, Sumertajaya IM. 2011. *Sidik Peubah Ganda Dengan Menggunakan SAS*. Bogor(ID): IPB Press.
- Rao JNK. 2003. *Small Area Estimation*. New York: John Wiley & Sons.
- Sud U.C., Chandra H. and Srivastava, A.K.. 2012. Crop Yield Estimation at District Level Using Improvement of Crop Statistics Scheme Data – An Application of Small Area Estimation Technique. *Journal of The Indian Society of Agricultural Statistics* 66(2): 321-326.