Karakteristik Pendugaan *Emperical Best Linear Unbiased Prediction* (EBLUP) Pada Pendugaan Area Kecil

M. Adi Sidauruk, Dian Kurniasari, Widiarti

Jurusan Matematika, FMIPA Universitas Lampung
E-mail: adisidauruk@yahoo.com

Abstrak. Pendugaan langsung pada parameter area kecil yang heterogen akan menghasilkan pendugaan yang tidak bias namun keragamannya sangat besar sehingga mengakibatkan pendugaan ini kurang valid. Untuk mengatasi masalah ini, Pendugaan pada area kecil dapat diduga dengan menggunakan pendugaan tidak langsung. Tujuannya adalah untuk menekan keragaman yang besar pada area kecil. Pendugaan tidak langsung pada area kecil memanfaatkan informasi dari area sekitarnya yang berhubungan dengan parameter yang menjadi perhatian. Ada beberapa metode pada pendugaan tidak langsung yaitu *Empirical Bayes* (EB), *Empirical Best Linear Unbiased Prediction* (EBLUP), dan *Hierarchical Bayes* (HB). Penelitian ini dilakukan pada data kontinu dengan metode EBLUP yang mengaplikasikan metode GLS sebagai metode penduga parameternya. Kajian secara teori menunjukkan bahwa penduga parameter yang dihasilkan oleh metode GLS merupakan penduga tak bias dengan varians minimum. Hasil yang sama juga diperoleh secara empiris melalui simulasi dengan *software R.2.10.1*.

Kata Kunci: Pendugaan Area Kecil, EBLUP, Tak-bias, Varians minim.

PENDAHULUAN

Pendugaan pada area kecil (*small area estimation*) merupakan salah satu upaya untuk menekan ragam yang besar pada area kecil yaitu dengan menggunakan pendugaan tidak langsung (*indirect estimation*) dengan memanfaatkan informasi dari area sekitarnya. Informasi yang diperoleh adalah informasi yang berhubungan dengan parameter.

Pendugaan menggunakan metode EBLUP pada data kontinu perlu dievaluasi karena penduga yang diperoleh pada area kecil merupakan penduga yang berbias namun memiliki ragam minimum. Tujuannya adalah untuk mendapatkan pendugaan yang efisien. Keakuratan penduga dapat diperoleh dengan cara mengukur *mean square error*-nya. Semakin kecil *mean square error* (MSE) suatu penduga maka penduga semakin akurat.

Rao (2003) mengemukakan bahwa suatu area disebut kecil apabila contoh yang diambil pada area tersebut tidak mencukupi untuk melakukan pendugaan langsung dengan hasil dugaan yang akurat. [3]

Pendugaan area kecil bertujuan untuk meningkatkan keakuratan penduga suatu parameter, yaitu dengan menggunakan pendugaan tidak langsung. Pendugaan tidak langsung dapat dilakukan dengan "meminjam kekuatan" atau memanfaatkan peubah-peubah tambahan dalam menduga parameter. Peubah pendukung ini berupa informasi tambahan yang didapatkan pada area lain dari survei yang sama, dari area yang sama pada survei yang terdahulu, atau peubah lain yang berhubungan dengan peubah yang menjadi perhatian pada area kecil.

Model area kecil merupakan model dasar dalam pendugaan area kecil. Dalam pendugaan area kecil terdapat dua jenis model dasar yang digunakan, yaitu *basic area level (Type A)* model dan *basic unit level (Type B)* model. [3]

Basic Area Level Model atau dapat disebut sebagai model berbasis area merupakan model yang didasarkan pada ketersediaan data pendukung yang hanya ada untuk level area tertentu, misalkan $x_i = (x_{1i}, x_{2i}, x_{3i}, \ldots, x_{pi})^T$ dengan parameter yang akan diduga adalah θ_i yang merupakan fungsi dari rata-rata peubah respon dan diasumsikan mempunyai keterkaitan dengan x_i. Data pendukung tersebut digunakan untuk membangun model

$$\theta_i = x_i^T\beta + b_i v_i, \quad (1)$$

dengan $i = 1, 2, \ldots, m$ dan $v_i \sim N(0, \sigma^2_v)$

Basic Area Level Model merupakan suatu model dimana data-data pendukung yang tersedia bersesuaian secara individu dengan data respon, misal $x_{ij} = (x_{ij1}, x_{ij2}, x_{ij3}, \ldots, x_{ijp})^T$ artinya untuk masing-masing anggota populasi j dalam masing-masing area kecil i, namun terkadang cukup dengan rata-rata populasi $\hat{x}_i$ diketahui saja. sehingga didapatkan suatu model regresi tersarang sebagai berikut :

$$y_{ii} = x_{ij}^T\beta + v_i + e_{ij}, \quad (2)$$

dengan $i = 1, 2, \ldots, m; j = 1, 2, \ldots, N_i$, dengan asumsi vi merupakan peubah acak yang berdistribusi $vi \sim N(0, \sigma^2_v)$ dan $e_{ij} = k\tilde{e}_{ij}$ dimana konstanta k diketahui dan $\tilde{e}_{ij}$ merupakan peubah acak saling bebas dari vi sehingga distribusi dari $\tilde{e}_{ij}$ adalah $\tilde{e}_{ij} \sim N(0, \sigma^2_e)$. Model dasar dalam pengembangan pendugaan area kecil didasarkan pada bentuk model linier campuran sebagai berikut :

$$y_i = x_i^T\beta + b_i v_i + e_i \quad (3)$$

Penduga terbaik (*best predictor*, BP) bagi $\theta_i = x_i^T\beta + v_i$ jika β dan A diketahui adalah

$$\hat{\theta}_i^{BP} = \hat{\theta}_i(y_i|A) = x_i^T\hat{\beta}_i + (1 - B_i)(y_i - x_i^T\hat{\beta}_i) \quad (4)$$

dengan $B_i = \frac{D_i}{A + D_i}$ untuk i = 1, 2, 3, …, m.

Jika A diketahui, β dapat diduga dengan metode kuadrat terkecil terboboti yaitu $\beta_i(A) = (X^T V^{-1} X)^{-1} X^T V^{-1} Y$ dan dengan mensubstitusi β oleh $\hat{\beta}_i$ pada $\hat{\theta}_i^{BP}$, maka diperoleh

$$\hat{\theta}_i^{BP} = \hat{\theta}_i(y_i|A) = x_i^T\hat{\beta}_i + (1 - B_i)(y_i - x_i^T\hat{\beta}_i)$$

$$\hat{\theta}_i^{BP} = \hat{\theta}_i(y_i|A) = (1 - B_i)y_i + B_i x_i^T\hat{\beta}_i$$

Penduga BLUP yang diperoleh dengan cara terlebih dahulu menduga komponen ragamnya. Kemudian mensubstitusi β oleh $\hat{\beta}$ dan A oleh $\hat{A}$ sehingga disebut sebagai prediksi tak-bias linear terbaik empirik (*Empirical Best Linear Unbiased Prediction*/EBLUP).

Generalisasi Kuadrat Terkecil

Perhatikan model linear

$$Y = X\beta + \varepsilon \quad (5)$$

diasumsikan matriks kovariansnya $\Sigma = \sigma^2\Delta(\sigma^2 < \infty)$ dengan σ^2 adalah parameter yang tidak diketahui nilainya dan Δ adalah matriks definit positif nxn dengan trase matriks sama dengan n. Jika suatu matriks Q adalah simetrik definit positif maka Q nonsingular atau Q^{-1} ada, dan karena itu ada matriks nxn nonsingular (misal P) sedemikian rupa sehingga

$$P'P = Q^{-1}$$

Matriks Δ adalah simetriks dan definit positif sehingga non-singular, karena itu ada suatu matriks nxn nonsingular P sehingga $P'P = \Delta^{-1}$. Pada model linear kalikan kedua ruas dengan matriks P ini :

$$PY = PX\beta + P\varepsilon \quad (6)$$

Penerapan metode kuadrat terkecil pada model di atas akan menghasilkan persamaan normal sebagai berikut :

$P'PPY = X'P'PXB$ dengan B adalah penduga kuadrat terkecil untuk β berdasarkan model di atas. Karena $X'P'PX$ adalah matriks definit positif jika X mempunyai peringkat kolom penuh *(full column rank)* sehingga $X'P'PX$ adalah nonsingular dan $P'P = \Delta^{-1}$ maka solusi persamaannya adalah

$$B = (X'P'PX)^{-1}X'P'PY$$

Atau

$$B = (X'\Delta^{-1}X)^{-1}X'\Delta^{-1}Y$$

persamaan terakhir ini dinamakan penduga kuadrat terkecil umum (*Generalized Least Squares*) untuk β selanjutnya disingkat dengan GLS. [4]

Pendugaan Parameter

Pendugaan parameter adalah proses untuk menduga atau menaksir parameter populasi yang tidak diketahui berdasarkan informasi dari sampel. Menurut Hoog dan

Craig (1995), kriteria penduga yang baik adalah takbias, varians minimum, konsisten, statistik cukup dan kelengkapan. [1]

Berikut ini hanya akan dibahas dua kriteria penduga yang baik, yaitu takbias dan varians minimum karena dianggap sudah cukup untuk melihat suatu penduga yang baik.

1. Takbias. Suatu statistik dikatakan penduga tidak bias dari parameter θ apabila nilai harapan penduga sama dengan parameter θ, sebaliknya jika nilai harapan statistik tersebut tidak sama dengan parameter θ maka disebut penduga θ yang berbias.

2. Varians Minimum. Suatu penduga $U(X)$ dikatakan mempunyai varians minimum apabila penduga tersebut memiliki varians yang kecil. Apabila terdapat lebih dari satu penduga, penduga yang efisiens adalah penduga yang memiliki varians terkecil.

Keakuratan suatu penduga menunjukkan tentang seberapa jauh penyimpangan nilai dugaan dari nilai parameter sebenarnya. Keakuratan suatu penduga umumnya dievaluasi berdasarkan nilai kuadrat galat / KTG (*mean square error* / MSE), yaitu MSE($\hat{\theta}$) = E(θ-$\hat{\theta}$) atau berdasarkan nilai akar kuadrat tengah galat / AKTG (*root mean square error* / RMSE), yaitu sebagai berikut

$$\mathrm{RMSE}(\hat{\theta}) = \sqrt{\mathrm{MSE}(\hat{\theta})}$$

$$= \sqrt{\mathrm{E}(\theta - \hat{\theta})2}$$

MSE ($\hat{\theta}^{BLUP}$) = $g_{1i}(A) + g_{2i}(A) = AD_i/(A+D_i) + (D_i)^2/(A+D_i)[X_i^t(X^tV^{-1}X)^{-1}X_i]$

METODE PENELITIAN

Data yang digunakan dalam penelitian ini merupakan data yang dibangkitkan dari simulasi dengan menggunakan *software R.2.10.1* dan sebaran datanya berdistribusi normal.

Adapun langkah- langkah yang dilakukan pada penelitian ini dalam mengkaji karakteristik penduga EBLUP pada penduga area kecil adalah sebagai berikut:

a. Menentukan nilai duga dari EBLUP dengan menggunakan metode GLS
b. Menentukan MSE dari nilai duga EBLUP
c. Mendapatkan karakteristik penduga EBLUP.

Langkah-langkah dalam menduga karakteristik penduga EBLUP dengan menggunakan simulasi, yaitu:

1. Membangkitkan peubah acak $x_i = (x_1, x_2,\ldots,x_{10})$ sebagai variabel peubah penyerta bagi variabel respon y_i dimana $x_i \sim N(10,1)$
2. Membangkitkan data $e_i = (e_1, e_2, \ldots, e_{10})$ sebagai sampling *error* dengan distribusi menyebar normal dimana nilai tengah dan varians bagi e_i yaitu 0 dan 0.5
3. Menetapkan nilai β
4. Membangkitkan data θ_i dengan iterasi 1000
5. Membangkitkan data y_i dengan iterasi 1000
6. Mendapatkan nilai $\hat{\beta}_i$ dengan iterasi 1000
7. Mendapatkan $\hat{\theta}_i^{EBLUP}$
8. Mengestimasi $\hat{\theta}_i^{EBLUP}$ sehingga memperoleh MSE dari $\hat{\theta}_i^{EBLUP}$.

HASIL DAN PEMBAHASAN

Penelitian ini menggunakan *software R 2.10.1* yang didesain untuk memperoleh penduga parameter β dan *mean square error* dari θ_i^{EBLUP} dengan metode *Emperical Best Linear Unbiased Predictions* (EBLUP). Simulasi ini menetapkan parameter β yang berbeda yaitu antara $0 \leq \beta \leq 1$ dengan *increase* 0.25 yang bertujuan untuk melihat konsintensi *mean square error*

(MSE/kuadrat tengah galat) parameter dari θ_i^{EBLUP}.

Karakteristik Penduga *Generalized Least Squares*

Misalkan Ω adalah matriks simetris definit positif. Faktor dari matriks ini dituliskan sebagai berikut

$\Omega = C\Delta C'$

C adalah karakteristik vektor Ω dan karakteristik akarnya adalah array dalam diagonal matriks Δ . Misalkan $\Delta^{1/2}$ adalah matriks diagonal dengan elemen diagonal ke- i yaitu $\sqrt{\lambda_i}$ dan $T = C\Delta^{1/2}$ sehingga $\Omega = TT'$.

Misalkan $P' = C\Delta^{-1/2}$ maka $\Omega^{-1} = P'P$. P dikalikan pada kedua ruas model linear $Y = X\beta + \varepsilon$ sedemikian sehingga diperoleh

$PY = PX\beta + P\varepsilon$

atau

$y_* = X_*\beta + \varepsilon_*$

Varians dari ε_* adalah $E[\varepsilon_*\varepsilon'_*] = P\sigma^2\Omega P' = \sigma^2 I$

dimana Ω diketahui, y_* dan X_* adalah data observasi. Pada model klasik, kuadrat tengah kecil (ordinary least squares) sangat effisien, oleh karena itu

$$\hat{\beta} = (X'_*X_*)^{-1}X'_*y_*$$
$$= (X'P'PX)^{-1}X'P'Py$$
$$= (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y$$

ini adalah penduga efisien dari β yang merupakan penduga *generalized least squares* (GLS). Adapun karakteristik penduga *generalized least squares* (GLS) adalah sebagai berikut

Tak-bias

Jika $E[\varepsilon_* I X_*] = 0$, sehingga

$$E[\hat{\beta}] = E[(X'_*X_*)^{-1}X'_*y_*]$$
$$= \beta + E[(X'_*X_*)^{-1}X'_*\varepsilon_*]$$
$$= \beta$$

Mendekati distribusi normal dengan *mean* β dan varians

$$Var(\hat{\beta}) = \sigma^2(X'_*X_*)^{-1} = \sigma^2(X'\Omega^{-1}X)^{-1}$$

Varians minimum [2]

Penduga parameter EBLUP adalah $\hat{\beta}$ dan $\hat{\theta}^{EBLUP}$. Selanjutnya akan dibahas mengenai karakteristik penduga EBLUP yaitu ketakbiasan dan ragam minimum. Pada penelitian ini, y dan θ diasumsikan menyebar normal. Karena y dan θ memiliki model masing-masing. Sehingga distribusi penduga parameter EBLUP dapat diketahui berdasarkan kedua model tersebut. Berikut ini adalah cara mengetahui distribusi pada penduga parameter EBLUP:

Karakteristik Penduga bagi β

Ketakbiasan

$$E(\hat{\beta} - \beta) = E(\hat{\beta}) - E(\beta)$$
$$= E((X^T V^{-1}X)^{-1}X^T V^{-1} y) - \beta$$
$$= (X^T V^{-1}X)^{-1}X^T V^{-1} E(y) - \beta$$
$$= (X^T V^{-1}X)^{-1}X^T V^{-1}X\beta - \beta$$
$$= \beta - \beta$$

= 0 terbukti

Ragam minimum

$$\text{Var}(\hat{\beta}) = \text{Var}((X^T V^{-1}X)^{-1}X^T V^{-1} y)$$
$$= (X^T V^{-1}X)^{-1}X^T V^{-1} Var(y)$$
$$= (X^T V^{-1}X)^{-1}X^T V^{-1}V$$
$$= (X^T V^{-1}X)^{-1}X^T$$

Karena $\hat{\beta}$ merupakan fungsi linear bagi e_i dan v_i dengan distribusi masing- masing menyebar normal maka diperoleh distribusi parameter $\hat{\beta}$ sebagai berikut

$$\hat{\beta} \sim N(\beta, (X^T V^{-1}X)^{-1}X^T)$$

Sebelumnya telah ditunjukkan bahwa $\hat{\beta}$ adalah penduga tak-bias bagi β. Oleh karena itu, akan ditunjukkan bahwa keragaman dari penduga β merupakan ragam minimum dengan menggunakan teorema Cramer- Rao berikut ini:

Pertama terlebih dahulu menghitung HB parameter penduga β

$$\left\{\frac{d}{d\beta}E[\text{w}(\text{x1}, \text{x2}, \ldots, \text{xn}]\right\}^2 = \left\{\frac{d}{d\beta}E[\hat{\beta}]\right\}^2$$
$$= \left\{\frac{d}{d\beta}E[\beta]\right\}^2$$
$$= 1$$

$$\frac{d}{d\beta}\ln f(\hat{\beta} \, I \, \beta)$$
$$= \frac{d}{d\beta}\ln\left[2n^{-1/2}\exp\left(-\frac{1}{2}\left(\frac{\hat{\beta}-\beta}{\sqrt{Var(\hat{\beta})}}\right)^2\right)\right]$$

$$= \frac{d}{d\beta}\left[\frac{-1}{2}\ln 2\text{n} - \exp\left(-\frac{1}{2}\left(\frac{\hat{\beta}-\beta}{\sqrt{Var(\hat{\beta})}}\right)^2\right)\right]$$

$$= 0 - \left(\frac{\hat{\beta} - \beta}{\sqrt{Var(\hat{\beta})}}\right)\frac{-1}{\sqrt{Var(\hat{\beta})}}$$

$$= \frac{\hat{\beta} - \beta}{Var(\hat{\beta})}$$

$$E\left[\frac{d}{d\beta} \ln f(\hat{\beta} \mid \beta)\right]^2 = E\left[\left(\frac{\hat{\beta} - \beta}{Var(\hat{\beta})}\right)\right]^2$$

$$= \frac{1}{Var^2(\hat{\beta})} E[(\hat{\beta} - \beta)]^2$$

$$= \frac{1}{Var^2(\hat{\beta})} Var(\hat{\beta})$$

$$= \frac{1}{Var(\hat{\beta})}$$

Sehingga diperoleh HB parameter $\hat{\beta}$ adalah $\frac{1}{\frac{1}{Var(\hat{\beta})}}$ atau sama dengan $var(\hat{\beta})$.

Karena $var(\hat{\beta})$ = HB maka $\hat{\beta}$ merupakan penduga tak-bias linear dengan varians minimum.

Hasil Simulasi Untuk Penduga Parameter $\boldsymbol{\beta}$

Simulasi pendugaan parameter β dengan metode EBLUP untuk iterasi 1000 dengan parameter set β = {0, 0.25, 0.5, 0.75, 1, 1.25, 1.75, 2} dan banyaknya pengamatan m = 10 menghasilkan 1000 penduga parameter β. Selanjutnya, akan dilakukan pembuktian ketakbiasan secara empirik dengan cara mencari rata-rata dari 1000 penduga yang diperoleh berdasarkan:

$$\bar{\hat{\beta}} = \frac{\hat{\beta}_{(1)} + \hat{\beta}_{(2)} + \hat{\beta}_{(3)} + \cdots + \hat{\beta}_{(1000)}}{\text{banyaknya iterasi}}$$

Tabel 4.1: Hasil Pendugaan Parameter $\hat{\boldsymbol{\beta}}$ pada beberapa nilai $\boldsymbol{\beta}$

$\boldsymbol{\beta}$	$\hat{\beta}$
0.00	0.03579911
0.25	0.2141868
0.50	0.5214076
0.75	0.7444331
1.00	0.9869544
1.25	1.235045
1.50	1.498268
1.75	1.765455
2.00	2.066544

Berdasarkan Tabel 4.1 dapat dilihat bahwa parameter $\hat{\beta}$ yang dihasilkan untuk masing-masing data set parameter β semakin mendekati nilai sebenarnya artinya berdasarkan pemeriksaan ketakbiasan diperoleh nilai $\bar{\hat{\beta}}$ mendekati nilai parameter β atau dapat disimpulkan bahwa bias parameter $\hat{\beta}$ terhadap data set parameter β sangat kecil dengan ragam minimum untuk setiap parameter β sama yaitu 0.001907544.

Hasil Simulasi Untuk Penduga Parameter $\boldsymbol{\theta^{EBLUP}}$

Mean Square Error dari setiap penduga parameter θ^{EBLUP} dapat diperoleh berdasarkan :

MSE($\hat{\theta}^{EBLUP}$) = E($\hat{\theta}^{EBLUP}$ - θ_i)^2
= var($\hat{\theta}^{EBLUP}$) +[Bias($\hat{\theta}^{EBLUP}$)]^2
= MSE($\hat{\theta}^{EBLUP}$) + E($\hat{\theta}^{EBLUP}$-$\hat{\theta}^{BLUP}$)^2

Prasad dan Rao dalam Rao (2003) menggunakan ekspansi deret taylor untuk menduga MSE (θ^{EBLUP}) sehingga diperoleh :

MSE(θ^{EBLUP}) = $g_{1i}(\hat{A}) + g_{2i}(\hat{A}) + g_{3i}(\hat{A})$

dengan $g_{3i}(\hat{A}) = \frac{2\hat{D}i^2}{m^2(\hat{A}+\hat{D}i)^3}\sum_{i=1}^{m}(\hat{A}+\hat{D}i)^2$

Setelah penduga parameter β disubstitusi ke dalam model θ_i^{EBLUP} = $x_i^t\beta + \left(\frac{\hat{A}}{\hat{A}+D_i}\right)(y_i - x_i^t\hat{\beta})$ maka diperoleh MSE($\hat{\theta}^{EBLUP}$) yaitu 0.8503704.

KESIMPULAN

Berdasarkan hasil dan pembahasan yang telah dilakukan, dapat disimpulkan bahwa:

a. Karakteristik parameter $\hat{\beta}$ yang diperoleh pada pendugaan area kecil merupakan penduga yang berbias namun bias yang ditimbulkan sangat kecil dengan varians minimum.
b. *Mean square error* (MSE) dari parameter $\hat{\theta}^{EBLUP}$ yang dihasilkan cukup kecil dengan menggunakan metode EBLUP pada pendugaan area kecil artinya nilai duga parameter

$\hat{\theta}^{EBLUP}$ cukup mendekati nilai parameter θ yang sebenarnya.

DAFTAR PUSTAKA

Hoog, R.V. dan Allen T. Craig. 1995. *Introduction to Mathematical Statistics*. Fifth edition. Princtice-Hall

Internasional Inc, New Jersey.

Greene, W. 1997. *Econometric Analysis*. Third edition. Prentice Hall, New Jersey

Rao, J. N. K. 2003. *Small Area Estimation*. New Jersey: John Willey & Sons, Inc.

Usman, M. dan Warsono. 2009. Teori Model Linear dan Aplikasinya. Sinar Baru Algensindo, Bandung.