BAB II

TINJAUAN PUSTAKA

1.1 Small Area Estimation

Suatu area disebut area kecil jika sampel yang diambil pada area tersebut tidak dapat mencukupi untuk melakukan pendugaan langsung dengan hasil dugaan yang akurat (Rao, 2003). Dewasa ini pendugaan area kecil menjadi sangat penting dalam analisis data survei karena adanya peningkatan permintaan untuk menghasilkan dugaan parameter yang cukup akurat dengan ukuran sampel yang kecil.

SAE merupakan suatu metode yang digunakan untuk memprediksi parameter dengan area kecil dengan memanfaatkan dari luar area, dalam area, dan dari luar survei (Longford, 2005).

Terdapat dua masalah pokok dalam pendugaan area kecil. Pertama adalah bagaimana meghasilkan suatu dugaan parameter yang cukup baik dengan ukuran sampel kecil pada suatu area kecil. Kedua yaitu bagaimana menduga *mean square error(MSE).* Solusi untuk masalah tersebut adalah dengan "meminjam informasi" dari dalam area, luar area, maupun luar survei (Pfefferman, 2002).

Metode SAE dapat digunakan untuk meningkatkan keakuratan pendugaandengan cara meningkatkan efisiensi penggunaan contoh melalui fungsi hubung (link function) antara penduga langsung dan pengaruh tetap dan pengaruh acak pada suatu area tertentu (Sadik*et al*, 2006).

Pendugaan parameter pada suatu area kecil dapat dilakukan dengan pendugaan secara langsung *(direct estimation)* maupun pendugaan secara tidak langsung *(indirect estimation)*. Hasil pendugaan langsung pada suatu area kecil merupakan pendugaan tak bias meskipun memiliki varian yang besar dikarenakan dugaannya diperoleh dari ukuran sampel yang kecil.

Berbagai survei umumnya dirancang untuk menduga parameter populasi untuk area yang besar, seperti level nasional, provinsi atau kabupaten/kota dimana pendugaan parameternya didasarka pada desain sampling. Jika akan menggunakan sampling area kecil, umumnya jumlah sampel kurang/tidak mencukupi. Oleh karena itu dikembangkan metode SAE untuk menduga parameter di suatu daerah dimana jumlah sampelnya berukuran kecil.

Terdapat dua ide utama yang digunakan dalam mengembangkan model pendugaan parameter area terkecil :

1. Model pengaruh tetap *(fixed effect model)* dimana asumsi bahwa keragaman dalam area kecil , variabel respon dapat diterangkan seluruhnya oleh hubungan keragaman yang bersesuaian pada informasi tambahan.
2. Model pengaruh acak area kecil *(random effect model)* dimana asumsi keragaman sepesifik area kecil tidak dapat diterangkan oleh informasi tambahan.

Gabungan dari model diatas membentuk model campuran *(mixed model)*. Karena variabel respon diasumsikan bahwa berdistribusi normal maka metode SAE dikembangkan merupakan bentuk kusus dari *General Linier Mixed Model* (GLMM). Terdapat dua dua tipe model pada metode SAE yakni model berbasis area level dan model berbasis unit level.

1. Model berbasis area level

Model berbasis area level merupakan model yang didasarkan pada ketersediaan data pendukung yang hanya dimiliki pada level area tertentu. Model ini diasumsikan bahwa θ_i berhubungn dengan variabel pelengkap X_i =(x_{1i},x_{2i}, ...,x_{pi})^T(Rao, 2003) melalui model linier berikut :

$$\theta_i = X_i^T \boldsymbol{\beta} + v_i \text{ ,} i = 1,\dots,m \quad (2.1)$$

Dimanaθ_i adalah penduga langsung dari survei untuk area kecil i, m adalah banyaknya area dengan $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_p)^T$merupakan vektor px1 koefisien regresi untuk variabel penyerta x_i dan v_i adalah pengaruh acak area kecil yang diasumsikan berdistribusi normal N(0,σ_v^2).

Dapat diketahui estimator θ_i dengan mengasumsikan bahwa model penduga langsung$\hat{\theta}_i$ telah tersedia, yaitu;

$$\hat{\theta}_i = \theta_i + e_i \text{ , } i = 1, \dots, m \quad (2.2)$$

Dengan e_i~ N(0,ψi) dan ψi di ketahui

Gabungan antara dua model (2.1) dan (2.2) akan menghasilkan persamaan model gabungan *(mixed model)* yang dikenal dengan model Fay-Herriot (Fay dan Herriot, 1979).

$$\hat{\theta}_i = \boldsymbol{x}_i^T \boldsymbol{\beta} + v_i + e_i, \ i=1, \dots, m \quad (2.3)$$

dimana x_i adalah vektor pxl variabel penyerta tingkat area v_i ~ N(0,σ_v^2) dan e_i~ N(0,ψi), dengan varian ψi, yang diketahui dari data, dimana v_i dan e_i saling bebas.

Dimana keragaman variabel respon di dalam area kecil di asumsikan dapat diterangkan oleh hubungan variabel respon dengan informasi tambahan (variabel penyerta) yang disebut dengan model pengaruh tetap *(fixed effect models)*. Selain itu terdapat komponen keragaman spesifik area kcil yang tidak bisa diterangkan oleh informasi tambahan (variabel penyerta), disebut dengan komponen pengaruh acak area kecil *(random effect)*. Gabungan dua asumsi tersebut membentuk model pengaruh acak campuran atau model linier campuran (Kurnia, 2009).

Saei dan Chambers (2003) mengemukakn dua ide utama dalam mengembangkan model SAE yaitu (1) asumsi bahwa keragaman didalam subpopulasi peubah respon dapat diterangkan seluruhnya oleh hubungan keragaman yang bersesuaian pada informasi tambahan, disebut model pengaruh tetap *(fixed effect)*, (2) asumsi keragaman spesifik subpoulasi tida dapat diterangkan oleh informasi tambahan dan merupakan pengaruh acak subpopulasi *(random effect)*. Gabungan dari kedua asumsi tersebut membentuk suatu pengaruh campuran *(mixed effect)*. Terjadi kelemahan jika model yang dibuat tidak menggambarkan kondisi wilayah/daerahyang sebenarnya.

2. Model berbasis unit level

   Model berbasis unit level merupakan suatu model dimana data-data pendukung yang tersedia bersesuaian secara individu dengan data respon, misal $x_i = x_{ij1}, x_{ij2}, \dots, x_{ijp})^T$, sehingga didapat suatu model regresi tersarang:

   $$y_{ij} = \boldsymbol{x}_{ij}^{\boldsymbol{T}}\boldsymbol{\beta} + v_i + e_{ij}\ ,\ i = 1, \dots, m \text{ dan } j = 1, \dots, n_i \tag{2.4}$$

   Dimana j adalah banyaknya IPM pada daerah ke-i dengan $v_i \sim N(0,\sigma_v^2)$ dan $e_i \sim N(0,\sigma_v^2)$.

Model yang digunakan dalam penelitian ini adalah model berbasis area, karena data penyerta yang digunakan merupakan data yang terdapat pada area tertentu yaitu pada level kabupaten/kota.

2.2 Metode *Empirical Best Linier Unbiased Prediction (EBLUP)*

EBLUP merupakan salah satu pendekatan yang digunakan untuk pendugaan area kecil, yang berawal dari ketidakmampuan *BLUP (Best Linier Unbiased Predictor)* dalam menduga komponen yang tidak diketahui. *BLUP* merupakan suatu pendugaan parameter yang meminimumkan *Mean Square Error* diantara kelas-kelas pendugaan parameter linier tak bias lainnya (Kurnia & Notodiputro, 2006). *BLUP* diasmsikan bahwa komponen ragam telah diketahui. Namun pada prakteknya, komponen ragam tidak diketahui. Oleh karena itu, diperlukan pendugaan terhadap komponen ragam tersebut melalui data contoh. Metode *EBLUP* mensubtitusi komponen ragam yang tidak diketahui ini dengan pendugaannya (Saei & Chambers, 2003).

Model Fay dan Harriot (1979) untuk dasar area level adalah :

$$y_i = \boldsymbol{x}_i^T \boldsymbol{\beta} + v_i + e_i, \; i = 1, \ldots, m \qquad (2.5)$$

$$= \theta_i + e_i$$

dimana x_i adalah vektor pxl variabel penyerta tingkat area v_i ~ N(0,σ_v^2) dan e_i~ N(0,ψi), dengan varian ψi, yang diketahui dari data, dimana v_i dan e_i saling bebas.

Penduga BLUP (Rao, 2003), pada θ_i, untuk asumsi σ_v^2 diketahui adalah :

$$\hat{\theta}_i^{\,BLUP} = \boldsymbol{x}_i^T \hat{\boldsymbol{\beta}} + v_i \qquad (2.6)$$

$$= \boldsymbol{x}_i^T\widehat{\boldsymbol{\beta}} + \gamma_i\,(\hat{\theta}_i - \boldsymbol{x}_i^T\widehat{\boldsymbol{\beta}})$$

dimana :

$\gamma_i = (\frac{\sigma_v^2}{\psi_i + \sigma_v^2})$

$\psi_i = MSE\,(\hat{\theta}_i) = \frac{s_i^{\,2}}{n_i}$, i = 1, ..., m

$\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}(\sigma_v^2)\left[\sum_{i=1}^{m}\frac{x_i x_i^T}{(\psi_i + \sigma_v^2)}\right]^{-1}\left[\sum_{i=1}^{m}\frac{x_i\,\hat{\theta}_i}{(\psi_i + \sigma_v^2)}\right]$

Dalam penduga BLUP terdapat nilai σ_v^2, karena pada metode BLUPdiasumikan bahwa σ_v^2 telah diketahui. Setelah nilai BLUPditemukana kemudian mencari nilai *MSE* untuk mengukur seberapa baik penduga BLUPyaitu dengan rumus :

$$MSE\,(\hat{\theta}_i^{BLUP}) = g1i(\sigma_v^2) + g2i(\sigma_v^2) \qquad (2.7)$$

Dalam prakteknya pengaruh acak*(random effect)* (σ_v^2) asumsi tidak diketahui. Terlebih dahulu harus diduga metode yang dapat digunakan untuk menduga varian pengaruh acak (σ_v^2) yaitu metode *Maximum Likelihood* (ML). Dengan metode ML di dapatkan iterasi Newton Rapson (Rao, 2003) sebagai berikut :

$$\sigma_v^{2\,(a+1)} = \sigma_v^{2\,(a)} + [\mathfrak{I}(\sigma_v^{2\,(a)})]^{-1} s(\widetilde{\boldsymbol{\beta}}^{(a)}, \sigma_v^{2\,(a)}) \qquad (2.8)$$

dengan :

$\mathfrak{I}(\sigma_v^2) = \frac{1}{2}\sum_{i=1}^{m}\frac{1}{(\sigma_v^2 + \psi_i)^2}$

$s(\widetilde{\boldsymbol{\beta}}^{(a)}, \sigma_v^{2\,(a)}) = -\frac{1}{2}\sum_{i=1}^{m}\frac{1}{(\sigma_v^2 + \psi_i)} + \frac{1}{2}\sum_{i=1}^{m}\frac{(\hat{\theta}_i - x_i^T\boldsymbol{\beta})^2}{(\sigma_v^2 + \psi_i)^2}$

Nilai $\sigma_v^{2(a+1)}$ dapat diambil sebagai penduga dari $\hat{\sigma}_v^2$ jika nilai $\sigma_v^{2(a+1)} = \sigma_v^2$. Kemudian nilai $\hat{\sigma}_v^2$diubah kedalam penduga BLUP, maka diperoleh hasil penduga baru yang disebut EBLUP (Rao, 2003), dengan rumus sebagai berikut :

$$\hat{\theta}_i^{EBLUP} = \boldsymbol{x}_i^T\hat{\boldsymbol{\beta}} + \hat{\gamma}_i(\hat{\theta}_i - \boldsymbol{x}_i^T\hat{\boldsymbol{\beta}}) \tag{2.9}$$

dimana:

$\hat{\gamma}_i = \left(\frac{\hat{\sigma}_v^2}{\psi_i + \hat{\sigma}_v^2}\right)$

$\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}\,(\hat{\sigma}_v^2) = [\sum_{i=1}^m \frac{x_i x_i^T}{(\psi_i + \hat{\sigma}_v^2)}]^{-1}[\sum_{i=1}^m \frac{x_i \hat{\theta}_i}{(\psi_i + \hat{\sigma}_v^2)}]$

Untuk mengukur seberapa baik penduga *EBLUP* maka dicari nilai *Mean Square Error MSE* dengan rumus :

$$MSE\,(\hat{\theta}_i^{EBLUP}) = g_{1i}(\hat{\sigma}_v^2) + g_{2i}(\hat{\sigma}_v^2) + 2g_{3i}(\hat{\sigma}_v^2) \tag{2.10}$$

dengan :

$g_{1i}(\hat{\sigma}_v^2) = \frac{\hat{\sigma}_v^2 \psi_i}{(\psi_i + \hat{\sigma}_v^2)} = \hat{\gamma}_i \psi_i$

$g_{2i}(\hat{\sigma}_v^2) = (1-\hat{\gamma}_i)^2 x_i^T [\sum_{i=1}^m \frac{x_i x_i^T}{(\psi_i + \hat{\sigma}_v^2)}]^{-1} x_i$

$g_{3i}(\hat{\sigma}_v^2) = \psi_i^2 (\psi_i + \hat{\sigma}_v^2)^{-3} \bar{V}(\hat{\sigma}_v^2)$

2.3. Korelasi Pearson Product Moment

Merupakan teknik pengujian korelasi yang dapat digunakan salah satunya adalah korelasi *pearson product moment* (*PPM*) , jika ingin mendapatkan nilai kesalahan terkecil. Korelasi Pearson menyatakan ada atau tidaknya hubungan yang signifikan antara variabel X dan Y. Nilai korelasi pearson antara X dan Y biasanya dilambangkan dengan r_{xy}. Asumsi yang harus dipenuhi dalam menggunakan korelasi *pearson* yakni setiap variabel berdistribusi normal, variabel yang dihubungkan memiliki hubungan yang linier, data dipilih secara acak *(random)*, data yang dihubungkan memiliki pasangan yang sama dan dari subjek yang sama pula dan variabel yang dihubungkan adalah data interval atau rasio (Usman dan Akbar, 2008).

Dengan uji kriteria dan hipotesis *pearson* sebagai berikut :

i. Hipotesis

$H_0: r_{xy} = 0$ (Tidak terdapat hubungan yang signifikan antara variabel X dengan nilai IPM dengan penduga langsung).

$H_0: r_{xy} \neq 0$ (Terdapat hubungan yang signifikan antara variabel X dengan nilai IPM dengan penduga langsung).

ii. Statistik uji:

Nilai korelasi *pearson* dirumuskan sebagai berikut :

$$r_{xy} = \frac{m\sum_{i=1}^{m} X_i Y_i - \left(\sum_{i=1}^{m} X_i\right)\left(\sum_{i=1}^{m} Y_i\right)}{\sqrt{\left\{m\sum_{i=1}^{m} X_i^2 - \left(\sum_{i=1}^{m} X_i\right)^2\right\}\left\{m\sum_{i=1}^{m} Y_i^2 - \left(\sum_{i=1}^{m} Y_i\right)^2\right\}}} \qquad (2.2)$$

dimana :

r_{xy} = Koefisien korelasi

m = Ukuran sampel

$\sum_{i=1}^{m} X_i$ = Jumlah pengamatan X

$\sum_{i=1}^{m} Y_i$ = Jumlah pengamatan

iii. Kriteria Pengujian

Tolak H_0 jika nilai sig. < α atau nilai $|r_{xy}|$ ≥ r tabel. Menurut Hasan (2005), nilai koefisien korelasi dapat diinterpretasikan sebagai berikut :

Tabel 2.1 Interpretasi korelasi X dan Y (r_{xy})

$\|r_{xy}\|$	Interpretasi
0	Tidak ada korelasi
$0 < r_{xy} \le 0{,}20$	Korelasi lemah sekali
$0{,}20 < r_{xy} \le 0{,}40$	Korelasi lemah
$0{,}40 < r_{xy} \le 0{,}70$	Cukup berkorelasi
$0{,}70 < r_{xy} \le 0{,}90$	Korelasi kuat
$0{,}90 < r_{xy} \le 1$	Korelasi kuat sekali
1	Sempurna

2.4. Uji Normalitas

Uji Normalitas dapat digunakan dengan Uji *Kolmogrov-smirnov*, pada pengujian *Kolmogrov-Smirnov* bahwa diasumsikan varibaelnya berdistribusi kontinu, maka data yang digunakan dalam uji ini tidak dapat diukur dengan skala ordinal. Uji *Kolmogrov-Smirnov* memiliki prinsip yaitumenghitung selisih absolut di masing-masing interval kelas. (Daniel, 1989). Uji *Kolmogrov-Smirnov* dapat dilakukan dengan hipotesis, statistik uji dan kriteria uji sebagai berikut :

i. Hipotesis :

$H_0: F(\hat{\theta}) = F_0(\hat{\theta})$ = Data berdistribusi normal.

$H_1: F(\hat{\theta}) \neq F_0(\hat{\theta})$ = Data tidak berdistribusi normal.

ii. Statistik Uji : α = 5%.

iii. Kriteria Uji : Tolak H_0 jika sig < α.

2.5. Indeks Pembangunan Manusia

IPM adalah pengukuran perbandingan dari harapan hidup, melek huruf, pendidikan dan standar hidup untuk semua negara di seluruh dunia. IPM digunakan untuk mengklasifikasikan apakah sebuah negara termasuk dalam kategori negara maju, negara berkembang atau negara tertinggal dan juga untuk mengukur pengaruh dari kebijaksanaan ekonomi terhadap kualitas hidup. Sedangkan menurut Badan Pusat Statistik IPM menjelaskan bagaimana penduduk dapat mengakses hasil pembangunan dalam memperoleh pendapatan, kesehatan, pendidikan, dan sebagainya. IPM diperkenalkan oleh *United Nations Development Programme* (UNDP) pada tahun 1990 dan dipublikasikan secara berkala dalam laporan tahunan *Human Development Report* (HDR). Pada tahun 2014 IPM mengalami dua penyempurnaan yaitu yang pertama mengganti tahun dasar Produk Nasional Bruto (PNB)per kapita dari tahun 2005 menjadi 2011. Yang kedua merubah metode agregasi indeks pendidikan dari rata-rata geometrik menjadi rata-rata aritmatik.

a. Angka Harapan Hidup

   Angka Harapan Hidup merupakan alat untuk mengevaluasi kinerja pemerintah dalam meningkatkan kesejahteraan penduduk pada umumnya, dan meningkatkan derajat kesehatan pada khususnya. Angka Harapan Hidup yang rendah di suatu daerah harus diikuti dengan program pembangunan kesehatan, dan program sosial lainnya termasuk kesehatan lingkungan, kecukupan gizi dan kalori termasuk program pemberantasan kemiskinan.

b. Rata-rata Lama Sekolah

Rata-rata Lama Sekolah merupakan jumlah tahun yang digunakan masyarakat untuk menjalani pendidikan formal. Diasumsikan bahwa dalam kondisi normal rata-rata lama sekolah suatu wilayah tidak akan turun. Penduduk yang dihitung dalam penghitungan rata-rata lama sekolah adalahpenduduk yang berusia 25 tahun ke atas.

c. Harapan Lama Sekolah

Harapan Lama Sekolah merupakan alat ukur yang digunakan untuk mengukur seberapa lamanya sekolah (dalam tahun) yang diharapkan dapat dinikmati oleh anak umur tertentu di masa mendatang. Di asumsikan bahwa peluang anak tersebut akan tetap bersekolah perjumlah penduduk untuk untuk umur yang sama saat ini. Angka Harapan Lama Sekolah dihitung mulai dari penduduk yang berusia 7 tahun ke atas. HLS dapat digunakan untuk mengetahui apakah setiap anak dapat merasakan pendidikan di setiap jenjang yang ditunjukan dalam bentuk lamanya sekolah (dalam tahun).

d. Pengeluaran per Kapita Disesuaikan

Pengeluaran per Kapita yang Disesuaikan ditentukan dari nilai pengeluaran per kapita dan paritas daya beli *(purchasing power parity-ppp).*

Rata-rata pengeluaran perkapita setahun diperoleh dari SUSENAS, dihitung dari level provinsi hingga level kab/kota. Rata-rata pengeluaran perkapita dibuat konstan/riil dengan tahun dasar 2012 = 100. Perhitungan paritas daya beli pada metode baru menggunakan 96 komoditas dimana 66 komoditas merupakan makanan dan sisanya komoditas nonmakanan. Metode penghitungan paritas daya beli menggunakan Metode Rao.