

II. TINJAUAN PUSTAKA

2.1 Penduga Area Kecil

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.

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. Keuntungan metode ini yaitu memiliki dugaan yang optimal, memperoleh model valid yang berasal dari data sampel, dan dapat menjelaskan berbagai macam model berdasarkan pada respon alami suatu kelompok dan kekelompokkan struktur data. Menurut Rao (2003), proses pendugaan pada suatu area atau subpopulasi terbagi menjadi dua, yaitu : pendugaan berbasis rancangan dan pendugaan berbasis model.

2.1.1 Pendugaan Berbasis Rancangan

Pendugaan ini merupakan penduga pada suatu area berdasarkan data contoh dari area itu sendiri. Proses pendugaan ini dapat menggunakan informasi tambahan untuk menduga parameter yang menjadi perhatian. Pendekatan klasik yang digunakan untuk menduga parameter area kecil didasarkan pada aplikasi model desain penarikan sampel yang menghasilkan metode pendugaan langsung dan diasumsikan tidak terjadi galat pengukuran.

2.1.2 Pendugaan Berbasis Model

Pendugaan pada metode berbasis model merupakan pendugaan suatu area dengan cara menghubungkan informasi pada area tersebut dengan area lain melalui model yang tepat. Hal ini berarti bahwa dugaan tersebut mencakup data dari area lain. Informasi yang digunakan diasumsikan memiliki hubungan dengan peubah yang menjadi perhatian. Tujuannya adalah untuk meningkatkan akurasi suatu penduga. Pendugaan parameter dan inferensianya yang berdasarkan pada informasi tambahan tersebut, dinamakan pendugaan tidak langsung atau pendugaan berbasis model (Rao, 2003). Metode pendugaan yang termasuk dalam penduga ini adalah metode EB, EBLUP, dan HB.

2.2 Model 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 (Rao 2003).

2.2.1 Basic Area Level (Type A) Model

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 $x_i = (x_{1i}, x_{2i}, x_{3i}, \dots, x_{pi})^T$ 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 $i = 1, 2, \dots, m$ dan $v_i \sim N(0, \sigma_v^2)$, sebagai pengaruh acak yang diasumsikan menyebar normal. Sedangkan b_i merupakan konstanta bernilai positif yang diketahui dan β adalah vektor koefisien regresi berukuran $p \times 1$. Kesimpulan mengenai θ_i , dapat diketahui dengan mengasumsikan bahwa model penduga langsung y_i telah tersedia, yaitu

$$y_i = \theta_i + e_i \quad (2)$$

dengan $i = 1, 2, \dots, m$ dan sampling error $e_i \sim N(0, \sigma_{ei}^2)$ dengan σ_{ei}^2 diketahui.

Dari kombinasi persamaan (1) dan (2) sehingga didapatkan model gabungan :

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

dengan $i = 1, 2, \dots, m$ dan dengan asumsi v_i dan e_i saling bebas. Rao (2003) menyatakan bahwa model tersebut merupakan bentuk khusus dari model linear campuran.

2.2.2 Basic Unit Level (Type B) Model

Merupakan suatu model dimana data-data pendukung yang tersedia bersesuaian secara individu dengan data respon, misal $x_{ij} = (x_{ij1}, x_{ij2}, x_{ij3}, \dots, x_{ijp})^T$ 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 :

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

dengan $i = 1, 2, \dots, m; j = 1, 2, \dots, N_i$, dengan asumsi v_i merupakan peubah acak yang berdistribusi $v_i \sim N(0, \sigma_v^2)$ dan $e_{ij} = k\tilde{e}_{ij}$ dimana konstanta k diketahui dan \tilde{e}_{ij} merupakan peubah acak saling bebas dari v_i sehingga distribusi dari \tilde{e}_{ij} adalah $\tilde{e}_{ij} \sim N(0, \sigma_e^2)$.

Model Fay-Herriot adalah model yang banyak dipakai dalam pendugaan area kecil dan merupakan model campuran linier. Fay dan Herriot menggunakan model dua level berikut untuk menduga pendapatan perkapita untuk area kecil di Amerika Serikat dengan populasi kurang dari 1000.

$$\text{Level 1 : } y_i | \theta_i \sim N(\theta_i, D_i)$$

$$\text{Level 2 : } \theta_i \sim N(x_i^t \beta, A)$$

Model dua level di atas dapat dituliskan sebagai model linear campuran sebagai berikut :

$$y_i = \theta_i + e_i = x_i^t \beta + v_i + e_i$$

Dengan $i = 1, \dots, m$, $v_i \sim N(0, A)$ dan $e_i \sim N(0, D_i)$.

Pengaruh acak area $v_i \sim N(0, A)$ digunakan untuk menghubungkan rata-rata area kecil θ_i dengan vektor peubah penyerta x_i yang sering diperoleh dari data sensus. Parameter β dan A umumnya tidak diketahui dan diduga dari sebaran marginal y . Ragam contoh D_i biasanya diasumsikan diketahui.

2.3 Metode *Empirical Best Linear Unbiased Predictions* (EBLUP)

Asumsi dasar dalam pengembangan untuk model pendugaan area kecil tersebut adalah keragaman di dalam area kecil peubah respon dapat diterangkan oleh hubungan keragaman yang bersesuaian pada informasi tambahan yang disebut pengaruh tetap. Asumsi yang lainnya yaitu bahwa keragaman spesifik area kecil tidak dapat diterangkan oleh informasi tambahan dan merupakan pengaruh acak area kecil. Gabungan dari dua asumsi tersebut membentuk model pengaruh campuran. Salah satu sifat yang menarik dalam model campuran adalah kemampuan dalam hal menduga kombinasi linear dari pengaruh tetap dan pengaruh acak. Henderson mengembangkan teknik penyelesaian model pengaruh campuran untuk memperoleh prediksi tak-bias linear terbaik (*best linear unbiased prediction* / BLUP). Menurut Rao (2003), BLUP merupakan suatu pendugaan parameter yang meminimumkan MSE diantara kelas-kelas pendugaan parameter linier tak bias lainnya. BLUP dihasilkan dengan asumsi bahwa komponen ragam diketahui. Namun faktanya, komponen ragam sulit bahkan tidak diketahui. Oleh karena itu, diperlukan pendugaan terhadap komponen ragam tersebut melalui data

sampel. Model dasar dalam pengembangan pendugaan area kecil didasarkan pada bentuk model linier campuran sebagai berikut :

$$y_i = x_i\beta + v_i + e_i \quad (4)$$

Dimana

y_i : nilai pendugaan langsung berdasarkan rancangan survei

x_i : variabel predictor yang elemen-elemennya diketahui.

β : vektor parameter bersifat *fixed* berukuran $px1$ yang tidak diketahui.

v_i : pengaruh acak area kecil dengan asumsi $v_i \sim N(0, \sigma_v^2)$ dimana $\sigma_v^2 = A$ dan biasanya tidak diketahui

e_i : vektor random *error* yang tidak terobservasi dengan asumsi $e_i \sim N(0, \sigma_{ei}^2)$ dimana $\sigma_{ei}^2 = D_i$ biasanya diasumsikan diketahui.

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

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

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

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).

Matriks *Expectations dan Variance Covariance* (VCV)

Secara umum ekspektasi dari y adalah

$$E \begin{pmatrix} y_i \\ v_i \\ e_i \end{pmatrix} = \begin{pmatrix} x_i \beta_i \\ 0 \\ 0 \end{pmatrix}$$

dan dikenal juga sebagai momen pertama. Momen kedua menggambarkan struktur *variance-covariance* dari y_i :

$$V(y_i) = A + D_i$$

diperoleh dari

$$V \begin{pmatrix} v_i \\ e_i \end{pmatrix} = \begin{pmatrix} A & 0 \\ 0 & D_i \end{pmatrix}$$

di mana D_i adalah matriks dispersi untuk efek *random* selain *error* dan A adalah matriks dispersi dari *error*, yang keduanya adalah matriks persegi umum diasumsikan untuk menjadi *non-singular* dan definit positif, dengan asumsi elemen-elemen diketahui.

2.4 Fungsi Pembangkit Momen

Momen dapat diperoleh melalui besaran lainnya, yang dinamakan fungsi pembangkit momen. Sehingga fungsi pembangkit momen merupakan sebuah fungsi yang dapat menghasilkan momen- momen. Selain itu, penentuan distribusi baru peubah acak yang baru merupakan kegunaan lain dari fungsi pembangkit momen.

2.4.1 Peubah Acak Tunggal

Misalkan X adalah peubah acak dengan *c.d.f* F_x dan *m.g.f* X dilambangkan dengan $M_x(t)$, yaitu

$$M_x(t) = E(e^{tx}),$$

Dengan syarat nilai harapannya ada untuk t di sekitar nilai nol. Yaitu terdapat $h > 0$ sedemikian rupa sehingga untuk semua t dalam $-h < t < h$, $E(e^{tx})$ ada. Jika nilai harapannya tidak ada di sekitar nol maka dikatakan bahwa *m.g.f* tidak ada.

Secara eksplisit dituliskan *m.g.f* X sebagai berikut

$$M_x(t) = \int_{-\infty}^{\infty} e^{tx} f_x(x) dx, \quad \text{jika } X \text{ kontinu,}$$

atau

$$M_x(t) = \sum_x e^{tx} P(X = x), \quad \text{jika } X \text{ diskrit.}$$

2.4.2 Peubah Acak Multipeubah/ Multivariat

Fungsi pembangkit momen untuk peubah acak multipeubah. Pertimbangkan peubah acak y_{nx1} . Fungsi pembangkit momen dari y dilambangkan dengan $m_y(\cdot)$ didefinisikan dengan

$$M_y(t) = E(e^{t'y}),$$

Jika dan hanya jika nilai harapan terdefinisi pada $-h < t < h$, $i = 1, 2, \dots, n$ untuk beberapa nilai $h > 0$. Jika nilai harapan tidak ada maka Y tidak mempunyai fungsi kepekatan peluang

2.5 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. 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 efisien adalah penduga yang memiliki varians terkecil.

2.6 Penduga Varians Minimum Seragam Takbias Linier

Misal Y_1, Y_2, \dots, Y_n adalah peubah acak dengan fungsi distribusi kumulatif $F_Y(\cdot | \gamma)$ dengan $\gamma \in \Omega$. misal $s(\gamma)$ sebagai fungsi γ dan diketahui, misalkan

dilambangkan dengan θ atau $\theta = s(\gamma)$. Jika $\hat{\theta}$ adalah fungsi linier Y_i , dan jika dalam semua kelas fungsi linier Y_i (dan untuk semua nilai $\gamma \in \Omega$), adalah takbias dan mempunyai varians terkecil dari semua penduga takbias dalam kelas fungsi linier Y_i maka $\hat{\theta}$ didefinisikan sebagai terbaik seragam (*uniformly best*) atau varians minimum, penduga linier takbias θ .

2.7 Generalisasi Kuadrat Terkecil (*Generalized Least Squares*)

Perhatikan model linear

$$Y = X\beta + \varepsilon$$

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 simetris 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 :

$$PY = PX\beta + P\varepsilon$$

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. (Usman, M dan Warsono, 2009)

2.7.1 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 efisien, oleh karena itu

$$\begin{aligned}\hat{\beta} &= (X_*' X_*)^{-1} X_*' y_* \\ &= (X_*' P' P X_*)^{-1} X_*' P' P y \\ &= (X_*' \Omega^{-1} X_*)^{-1} X_*' \Omega^{-1} y\end{aligned}$$

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

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

- Konsisten

Jika $\text{plim} \left(\frac{1}{n} X_*' X_* \right) = Q_*$, dimana Q_* adalah matriks berhingga definit positif

- Mendekati distribusi normal dengan *mean* β dan varians

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

- Varians minimum (Greene, W)

2.8 Teorema Gauss-Markov

Perhatikan model linear umum

$$Y = X\beta + \varepsilon \text{ dengan } E(\varepsilon) = 0 \text{ dan } \text{Cov}(\varepsilon) = \sigma^2 I,$$

Adalah model dengan peringkat penuh dan $\Omega = \{(\beta, \sigma^2) | \beta \in E_p, \sigma^2 > 0\}$.

penduga kuadrat terkecil $l'\beta$ (l adalah vektor konstan $p \times 1$) diberikan oleh $l'\hat{\beta}$

(yaitu $l'\hat{\beta} = l'X^{-1}Y$) dan ini merupakan penduga dengan varians minimum seragam linear takbias untuk parameter $l'\beta$ dengan $\hat{\beta} = (X'X)^{-1}X'Y$ (Usman, M. dan Warsono, 2009).

Teorema Cramer-Rao Lower Bound

Perhatikan peubah acak (x_1, x_2, \dots, x_n) dengan fungsi kepekatan peluang $f(x_1, x_2, \dots, x_n | \theta)$ dan $\hat{\theta}$ adalah penduga tak bias bagi θ ($\hat{\theta}$ merupakan fungsi dari $x = w(x)$). Misalkan bahwa $f(x_1, x_2, \dots, x_n | \theta)$ memiliki sifat berikut

$$\frac{\partial}{\partial \theta} \int \dots \int h(x_1, \dots, x_n) f(x_1, x_2, \dots, x_n | \theta) dx_1 dx_2 \dots dx_n$$

Maka

$$\text{Var}(\theta) = \text{Var}[w(x_1, x_2, \dots, x_n)] \geq \frac{\left\{ \frac{d}{d\theta} E[w(x_1, x_2, \dots, x_n)] \right\}^2}{E\left[\frac{d}{d\theta} \ln f(x_1, x_2, \dots, x_n | \theta) \right]^2}$$

dimana $\frac{\left\{ \frac{d}{d\theta} E[w(x_1, x_2, \dots, x_n)] \right\}^2}{E\left[\frac{d}{d\theta} \ln f(x_1, x_2, \dots, x_n | \theta) \right]^2}$ disebut sebagai batas bawah atau had bawah (HB)

Corollary

Kasus khusus dari teorema Cramer- Rao jika x_1, x_2, \dots, x_n iid (*independen identic distributed*) maka

$$\text{Var}[w(x_1, x_2, \dots, x_n)] \geq \frac{\left\{ \frac{d}{d\theta} E[w(x_1, x_2, \dots, x_n)] \right\}^2}{n E\left[\frac{d}{d\theta} \ln f(x_1, x_2, \dots, x_n | \theta) \right]^2}$$

2.9 Mean Square Error (MSE)

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

$$\text{MSE}(\hat{\theta}) = E(\theta - \hat{\theta})^2$$

Atau berdasarkan nilai akar kuadrat tengah galat / AKTG (*root mean square error* / RMSE), yaitu sebagai berikut

$$\begin{aligned} \text{RMSE}(\hat{\theta}) &= \sqrt{\text{MSE}(\hat{\theta})} \\ &= \sqrt{E(\theta - \hat{\theta})^2} \end{aligned}$$

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

Berdasarkan definisi EBLUP, penduga EBLUP dapat diperoleh dengan cara mensubstitusi komponen ragam \hat{A} dan $\hat{\beta}$ masing- masing ke komponen ragam yang tidak diketahui, yaitu A dan β . Maka akan diperoleh penduga EBLUP, yaitu:

$$\begin{aligned} \hat{\theta}^{\text{EBLUP}} &= \hat{\theta}_i(y_i I \hat{A}) \\ &= x_i^t \beta + \left(\frac{\hat{A}}{\hat{A} + D_i} \right) (y_i - x_i^t \hat{\beta}) \end{aligned} \tag{6}$$

Dan MSE dari $\hat{\theta}^{\text{EBLUP}}$ adalah

$$\begin{aligned} \text{MSE}(\hat{\theta}^{\text{EBLUP}}) &= E(\hat{\theta}^{\text{EBLUP}} - \theta_i)^2 \\ &= \text{var}(\hat{\theta}^{\text{EBLUP}}) + [\text{Bias}(\hat{\theta}^{\text{EBLUP}})]^2 \\ &= \text{MSE}(\hat{\theta}^{\text{EBLUP}}) + E(\hat{\theta}^{\text{EBLUP}} - \hat{\theta}^{\text{BLUP}})^2 \end{aligned}$$

dengan menggunakan ekspansi deret Taylor untuk menduga MSE ($\hat{\theta}^{\text{EBLUP}}$)

sehingga diperoleh :

$$\text{MSE}(\hat{\theta}^{\text{EBLUP}}) = g_{1i}(\hat{A}) + g_{2i}(\hat{A}) + g_{3i}(\hat{A}) \quad (7)$$

$$\text{Dengan } g_{3i}(\hat{A}) = \frac{2\hat{D}i^2}{m^2(\hat{A} + \hat{D}i)^3} \sum_{i=1}^m (\hat{A} + \hat{D}i)^2 \text{ (Rao, 2003).}$$