Kebutuhan Rumah Sederhana di Kabupaten Jember dengan *Robust Small Area Estimation*

Simple House Needs in Jember with Robust Small Area Estimation

Frida Murtinasari*), Alfian Futuhul Hadi, Dian Anggraeni
Program Studi Magister Matematika, FMIPA Universitas Jember, Jember
*)Email: fnopiyanto@gmail.com

ABSTRACT

SAE (Small Area Estimation) is often used by researchers, especially statisticians to estimate parameters of a subpopulation which has a small sample size. Empirical Best Linear Unbiased Prediction (EBLUP) is one of the indirect estimation methods in Small Area Estimation. The presence of outliers in the data can not guarantee that these methods yield precise predictions . Robust regression is one approach that is used in the model Small Area Estimation. Robust approach in estimating such a small area known as the Robust Small Area Estimation. Robust Small Area Estimation divided into several approaches. It calls Maximum Likelihood and M-Estimation. From the result, Robust Small Area Estimation with M-Estimation has the smallest RMSE than others. The value is 1473.7 (with outliers) and 1279.6 (without outlier). In addition the research also indicated that REBLUP with M-Estimation more robust to outliers. It causes the RMSE value with EBLUP has five times to be large with only one outlier are included in the data analysis. As for the REBLUP method is relatively more stable RMSE results.

Keywords : SAE, EBLUP, Robust, Maximum Likelihood, M-Estimation

PENDAHULUAN

SAE (*Small Area Estimation*) merupakan salah satu metode dalam statistik yang sudah tidak asing lagi. SAE sering digunakan oleh para peneliti terutama statistisi untuk menduga parameter – parameter subpopulasi yang memiliki ukuran sampel kecil (Rao, 2003). Dalam SAE terdapat dua teknik pendugaan yaitu pendugaan langsung (*direct estimation*) dan pendugaan tidak langsung (*indirect estimation*). Rao (2003) menyatakan bahwa pendugaan sederhana area kecil yang didasarkan pada penerapan model desain penarikan contoh (*design-based*) disebut sebagai pendugaan langsung (*direct-estimation*). Pendugaan langsung tidak mampu memberikan ketelitian yang cukup bila ukuran sampel dalam area kecil yang menjadi perhatian sedikit/ berukuran kecil, sehingga statistik yang dihasilkan akan memiliki varian yang besar atau bahkan pendugaan tidak dapat dilakukan karena tidak terwakili dalam survey. Karena hal tersebut maka dikembangkan metode pendugaan tidak langsung agar ragam yang dihasilkan semakin kecil. Metode tidak langsung antara lain *Empirical Best Linear Unbiased Prediction* (EBLUP), *Empirical Bayes* (EB), dan *Hierarchical Bayes* (HB). Metode yang lebih umum digunakan adalah EB dan HB karena dapat digunakan untuk data kontinu, cacahan maupun biner. Metode ini memasukkan variabel penyerta (*model-based*) yang dihubungkan dengan area terkait sebagai informasi tambahan.

Empirical Best Linear Unbiased Prediction (EBLUP) merupakan salah satu metode pendugaan tak langsung dalam *Small Area Estimation*. Adanya outlier pada data tidak menjamin bahwa metode ini dapat menghasilkan prediksi yang tepat. Data yang memiliki outlier tidak dapat dibuang begitu saja karena akan mempengaruhi model prediksi serta menghasilkan estimasi yang kurang tepat (Mustikasari, 2012). Menurut Aunuddin (1989) regresi robust dianggap memiliki kelebihan karena dapat menanggulangi data outlier, sehingga mengurangi sifat bias pada penduga yang dihasilkan dan prediksinya menjadi lebih tepat. Pada tahun 2003, Rao menuangkan ide mengenai outlier pada *Small Area Estimation* dalam *Robust Small Area Estimation*. Kemudian Tobias Schoh melanjutkan riset tersebut dengan mengembangkan paket *Robust Small Area Estimation* dalam software R. *Robust Small Area Estimation* dapat diperoleh dengan beberapa pendekatan antara lain yaitu dengan *Maximum Likelihood* dan *M-Estimation*.

Munculnya outlier pada data dapat mengakibatkan prediksi yang dihasilkan oleh beberapa metode pendugaan dalam *Small Area Estimation* kurang tepat. Oleh karena itu dalam penelitian ini akan dibandingan sensitivitas ketiga metode pendugaan dalam *Small Area Estimation* yaitu EBLUP, REBLUP *Maximum Likelihood* maupun REBLUP *M-Estimation* pada saat ada data outlier didalamnya. Dalam kondisi tersebut dapat diamati apakah metode pendugaan dalam *Robust Small Area Estimation* lebih stabil atau konsisten terhadap outlier data.

Adapun permasalahan yang diangkat dalam penelitian ini adalah mengenai kebutuhan rumah sederhana. Di tingkat kabupaten terutama Kabupaten Jember, kebutuhan akan rumah terus meningkat sejalan dengan pertambahan jumlah penduduk sehingga subsidi Kredit Pemilikan Rumah(KPR) untuk perumahan juga mengalami peningkatan. Oleh karena itu pemerintah dituntut untuk lebih tepat sasaran dalam memberikan bantuan terhadap masyarakat. Disisi yang lain, data yang diperoleh untuk prediksi kebutuhan tersebut terbatas. Sehingga pada penelitian ini akan diduga kebutuhan rumah sederhana di kabupaten Jember menggunakan *Robust Small Area Estimation* .

Small Area Estimation Model

Menurut Rao (2003), proses pendugaan pada suatu area atau subpopulasi terbagi menjadi dua, yaitu :

a. Pendugaan Berbasis Rancangan (*design based*) . Pendugaan ini merupakan penduga pada suatu area berdasarkan data contoh dari area itu sendiri.
b. Pendugaan Berbasis Model (*Model Based*). Pendugaan pada metode berbasis model merupakan pendugaan suatu area dengan cara menghubungkan informasi pada area tersebut dengan area lain melalui model yang tepat.

Darsyah (2013) menyebutkan ada dua konsep pokok yang digunakan untuk mengembangkan *model small area*, yaitu

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

Gabungan antara kedua model tersebut membentuk model campuran (*mixed model*). Karena variabel respon diasumsikan berdistribusi normal maka pendugaan area kecil yang dikembangkan merupakan bentuk khusus dari *General Linear Mixed Model* (GLMM).

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.

a. 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 atau penyerta 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 atau penyerta tersebut digunakan untuk membangun model

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

dengan $i = 1,2,\ldots,m$ dan $v_i \sim N(0,\sigma_v^2)$ sebagai pengaruh acak yang diasumsikan menyebar normal. Sedangkan b_i merupakan konstanta bernilai positif yang diketahui dan β adalah vektor koefisien regresi berukuran p x 1. Estimator dari θ_i dapat diketahui dengan mengamsusikan bahwa model penduga langsung y_i telah tersedia, yaitu

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

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

Dari kombinasi persamaan (1) dan (2), didapatkan model linier campuran sebagai berikut :

$$\hat{\theta}_i - x_i^T\beta + b_i v_i + e_i \quad (3)$$

dengan $i = 1, 2, \ldots, m$ dan dengan asumsi v_i dan e_i saling bebas.

b. Basic Unit Level (Type B) Model

Merupakan suatu model dimana data-data pendukung yang tersedia bersesuaian secara individu dengan data respon, misal $\mathbf{x_{ij}} = (\mathbf{x_{ij1}}, \mathbf{x_{ij2}}, \mathbf{x_{ij3}}, \ldots, \mathbf{xi_{jp}})^T$ artinya untuk masing-masing anggota populasi j dalam masing-masing area kecil i, namun terkadang cukup dengan rata-rata populasi

x_i diketahui saja. sehingga didapatkan suatu model linier campuran sebagai berikut :

$$\hat{\theta}_{ij} = x_{ij}^T\beta + v_i + e_{ij} \quad (4)$$

dengan *i = 1, 2, ..., m; j = 1, 2, ..., Ni*, dengan asumsi v_i merupakan peubah acak yang berdistribusi $v_i \sim N(0, \sigma_{v_i}^2)$ dan $\tilde{e}_{ij}$ merupakan peubah acak saling bebas dari v_i sehingga distribusi dari adalah $\tilde{e}_{ij} \sim N(0, \sigma_e^2)$.

Metode *Empirical Best Linear Unbiased Predictions* (EBLUP) dalam *Small Area Estimation*

Asumsi dasar dalam pengembangan model pendugaan area kecil tersebut adalah keragaman di dalam area kecil, dimana peubah respon dapat diterangkan oleh hubungan keragaman yang bersesuaian dengan informasi tambahan yang disebut pengaruh tetap. Di sisi yang lain terdapat asumsi yaitu bahwa keragaman spesifik area kecil tidak dapat diterangkan oleh informasi tambahan yang 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.

Rao (2003) mengungkapkan bahwa model untuk pendugaan tidak langsung, yaitu $\hat{\theta}_i = x_i^T\beta + b_i v_i + e_i$, *i* = 1,...,*m* ,sebenarnya merupakan kasus khusus dari model linier campuran pada persamaan (1), dengan $Y_i = \hat{\theta}_i$, $\mathbf{X}_1 = \mathbf{x}_1^T$ dan $\mathbf{Z}_1 = \mathbf{b}_1$ serta $\mathbf{V}_1 = \mathbf{v}_i$, $e_i - e_i$, dan $\beta = (\beta_1, \ldots, \beta_p)^T$.

Secara umum matriks koragam dari bentuk persamaan linier campuran $y^p = \mathbf{x}^p\boldsymbol{\beta} + Z^p\mathbf{v} + \mathbf{e}^p$ bagi $\mathbf{v}$ dan $\mathbf{e}$ adalah G dan **R** . Dalam bentuk linier campuan ini dilakukan pendugaan terhadap kombinasi linier dari parameter yaitu $\mu = \mathbf{1}^T\boldsymbol{\beta} + \mathbf{m}^T\mathbf{v}$. Rao (2003) mengemukakan bahwa untuk δ tertentu yang diketahui maka penduga BLUP bagi μ adalah

$$\tilde{\mu}^H = t(\delta, y) = \mathbf{1}^T\tilde{\boldsymbol{\beta}} + \mathbf{m}^T\mathbf{v} = \mathbf{1}^T\tilde{\boldsymbol{\beta}} + \mathbf{m}^T\mathbf{G}\mathbf{Z}^T\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}) \quad (5)$$

dimana

$$\tilde{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}(\delta) = (\mathbf{x}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{x}^T\mathbf{V}^{-1}\mathbf{y}$$

$$\tilde{\mathbf{v}} = \tilde{\mathbf{v}}(\delta) = GZV^{-1}(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}),$$

dengan $\tilde{\boldsymbol{\beta}}$ merupakan matriks vektor koefisien regresi yang berukuran p x 1 dari $\boldsymbol{\mu}$.

Pada model linier campuran dalam *Small Area Estimation*, $\mathbf{G}_1 = \sigma v^2$, $\mathbf{R}_1 = \psi_i$ sehingga diperoleh $\mathbf{V}_1 = \psi_i + \sigma v^2 bi^2$ dan $\mu_i = \theta_i = z_i^T\beta + b_i v_i$. Oleh karena itu penduga BLUP bagi μ_i atau θ_i dapat dituliskan sebagai berikut :

$$\tilde{\theta}_i^H = t(\delta, y) = \mathbf{z}_i^T\tilde{\boldsymbol{\beta}} + \gamma_i(\bar{y}_i - \mathbf{z}_i^T\tilde{\boldsymbol{\beta}}) \quad (6)$$

dimana $\gamma_i = \sigma v^2 bi^2/(\psi_i + \sigma v^2 bi^2)$.

Sehingga diperoleh penduga terbaik (*best predictor*, BP) bagi $\hat{\theta}_i = x_i^T\beta + b_i v_i + e_i$ jika β dan *A* diketahui adalah sebagai berikut:

$$\hat{\theta}_i^{BP} = \hat{\theta}_i(y_i l\beta, A) = x_i^T\beta + (1 - B_i)(y_i - x_i^T\beta) \quad (7)$$

dengan

$B_i = \frac{D_i}{A + D_i}$ untuk *i* = 1,2,3,...m, A = $\sigma v^2 bi^2$ dan $D_i = \mathbf{R}_1 = \psi_i$.

Jika A diketahui, maka $\hat{\theta}_i^{EBLUP}$ dapat diduga dengan metode kuadrat terkecil terboboti yaitu dengan mensubstitusi β oleh $\hat{\beta}_i$ pada $\hat{\theta}_i^{EBLUP}$sehingga diperoleh

$$\begin{aligned}
\theta_i^{EBLUP} &= \theta_i(y_i l, A) = x_i^T\beta_i + (1 - B_i)(y_i - x_i^T\beta_i) \\
&= x_i^T\beta_i + (1 - B_i)y_i - (1 - B_i)x_i^T\beta_i \\
&= x_i^T\beta_i + y_i - B_i y_i - x_i^T\beta_i + B_i x_i^T\beta_i \\
&= (1 - B_i)y_i + B_i x_i^T\beta_i. \quad (8)
\end{aligned}$$

Penduga BLUP yang diperoleh dengan cara terlebih dahulu menduga komponen ragamnya, kemudian mensubstitusi β oleh $\hat{\beta}$ dan *A* oleh $\hat{A}$ yang disebut sebagai prediksi tak-bias linear terbaik empirik (*empirical best linear unbiased prediction* / EBLUP). Sedangkan bentuk *Mean Square Error*nya adalah

$$\text{MSE}(\hat{\theta}_i^{EBLUP}) = \text{E}(\hat{\theta}_i^{EBLUP} - \theta_i)^2.$$

$$= \frac{1}{i}\sum(\hat{\theta}_i^{EBLUP} - \theta_i)^2 \quad (9)$$

Meskipun metode EBLUP klasik berguna untuk memperkirakan area kecil yang efisien di bawah asumsi normalitas, akan tetapi hal tersebut dipengaruhi oleh adanya outlier dalam data. Terutama apabila parameter tidak dapat diperkirakan secara konsisten dengan adanya kontaminasi pada data. Dan apabila terdapat kontaminasi, metode apapun untuk menduga hal ini akan bias. Sehingga bisa dikatakan bahwa metode EBLUP klasik menjadi tidak efisien untuk menduga $\bar{\theta}$ (Schoch, 2012).

Terdapat dua metode yang digunakan untuk menduga parameter dalam Robust Small Area Estimation, yaitu dengan Robust EBLUP *Maximum Likelihood* dan Robust EBLUP M-*Estimation*.

a. REBLUP *Maximum Likelihood* (RML)

Schoch (2012) mengemukakan bahwa pada dasarnya pendekatan REBLUP dengan Maximum Likelihood memiliki tingkat keakuratan yang hampir sama dengan metode EBLUP yang juga menggunakan *Maximum Likelihood*. Yang membedakan REBLUP ML dan EBLUP adalah dipastikannya outlier yang terdapat pada data yang digunakan pada REBLUP serta adanya prediktor yang diasumsikan robust outlier pada β dan v.

Wels and Richardson (1997) menyatakan bahwa Maximum Likelihood akan memiliki bias yang besar apabila parameter tidak dapat diperkirakan secara konsisten sehingga estimator yang menggunakan Maximum Likelihood bisa menjadi sangat tidak efisien. Parameter θ yang akan diduga ($\bar{\theta}$) merupakan fungsi rata-rata yang dari peubah respon yaitu Y. Pada persamaan (5), Sinha dan Rao (2009) menggantikan $\hat{\beta}$ dan $\hat{v}$ dengan prediktor robust outlier yaitu $\hat{\beta}^{ROBUST}$ dan $\hat{v}^{ROBUST}$ untuk membuat prediktor yang lebih resisten terhadap outlier. Sehingga REBLUP ML θˆ adalah sebagai berikut :

$$\begin{aligned}\hat{\theta}_i^{REBLUP} &= \hat{\theta}_i(y_i, I, A) = x_i^T \hat{\beta}^{ROBUST}{}_i + (1-B_i)(y_i - x_i^T \hat{\beta}^{ROBUST}{}_i) \\ &= x_i^T \hat{\beta}^{ROBUST} + (1-B_i)y_i - (1-B_i)x_i^T \hat{\beta}^{ROBUST} \\ &= x_i^T \hat{\beta}^{ROBUST} + y_i - B_i y_i - x_i^T \hat{\beta}^{ROBUST} + B_i x_i^T \hat{\beta}^{ROBUST} \\ &= (1-B_i)y_i + B_i x_i^T \hat{\beta}^{ROBUST}\end{aligned} \tag{10}$$

dengan

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

Sedangkan bentuk dari *Mean Square Errornya* adalah sebagai berikut :

$$MSE(\hat{\theta}_i^{REBLUP}) = E(\hat{\theta}_i^{REBLUP} - \theta_i)^2.$$

$$= \frac{1}{i}\sum(\hat{\theta}_i^{EBLUP} - \theta_i)^2 \tag{11}$$

b. REBLUP M-Estimation

Prosedur robust digunakan untuk mengakomodasi adanya keanehan data, sekaligus meniadakan identifikasi adanya data outlier dan juga bersifat otomatis dalam menanggulangi data outlier (Aunuddin, 1989). Chen (2002) menyebutkan beberapa prosedur estimasi parameter dalam regresi robust, salah satu diantaranya adalah M-Estimation yang diperkenalkan Huber (1973) .

Myers (1990) menyatakan bahwa M-Estimation merupakan salah satu metode dalam regresi robust yang sering digunakan untuk mengestimasi parameter yang disebabkan oleh x-outlier dan memiliki breakdown point 1/n. M-estimation Huber melalui fungsi ψ (.) melibatkan pengkuadratan residual yang kecil seperti pada Ordinary Least Square tetapi memberikan residual yang besar sedemikian rupa untuk mengurangi pengaruh robust). Pada Robust Small Area Estimation, outlier terjadi pada pengaruh acak yaitu pada β dan v sehingga fungsi ψ (.) berada pada penduga robust untuk β dan v. Payam dan Ray (2013) menuliskan bentuk prediktor mean dari REBLUP dengan M-Estimation adalah sebagai berikut :

$$\begin{aligned}\hat{\theta}_i^{REBLUP\,MST} &= \hat{\theta}_i(y_i, I, A) = x_i^T \hat{\beta}^{\psi}{}_i + (1-B_i)(y_i - x_i^T \hat{\beta}^{\psi}{}_i) \\ &= x_i^T \hat{\beta}^{\psi} + (1-B_i)y_i - (1-B_i)x_i^T \hat{\beta}^{\psi} \\ &= x_i^T \hat{\beta}^{\psi} + y_i - B_i y_i - x_i^T \hat{\beta}^{\psi} + B x_i^T \hat{\beta}^{\psi} \\ &= (1-B)y_i + B x_i^T \hat{\beta}^{\psi}\end{aligned} \tag{12}$$

dengan $\hat{\beta}^{\psi}$ dan $\hat{v}^{\psi}$ merupakan penduga *robust* dari pengaruh tetap dan pengaruh acak model. Sedangkan bentuk *Mean Square Error*nya yaitu :

$$MSE(\hat{\theta}_i^{REBLUP\,MST}) = E(\hat{\theta}_i^{REBLUP\,MST} - \theta_i)^2.$$

$$= \frac{1}{i}\sum(\hat{\theta}_i^{REBLUP\,MST} - \theta_i)^2 \tag{13}$$

METODE

Sumber data yang digunakan dalam penelitian jumlah kebutuhan rumah sederhana di kabupaten Jember dengan Robust Small Area Estimation adalah data Survei Sosial Ekonomi Nasional (SUSENAS) BPS Tahun 2012 dan variabel penyerta berasal dari pendataan Potensi Desa yang dituliskan pada Jember Dalam Angka Tahun 2011. Variabel respon yang menjadi perhatian dalam penelitian ini adalah jumlah kebutuhan rumah sederhana pada masing-masing kecamatan di kabupaten Jember.

Adapun model small area yang digunakan pada penelitian ini adalah model small area dengan menggunakan pendekatan basic unit level model. Hal ini disebabkan karena data kebutuhan rumah sederhana bersesuaian sampai unit kecamatan. Sehingga dapat dituliskan untuk model regresi tersarang dari small area adalah sebagai berikut :

$$\hat{\bar{y}}_{ij} = \hat{\theta}_{ij} = x_{ij}^T\beta + v_i + e_{ij},$$

dengan $\hat{\theta}_{ij}$ merupakan nilai penduga kebutuhan rumah sederhana kecamatan di kabupaten Jember. Sedangkan x_i^T merupakan variabel penyerta berupa jumlah keluarga menengah ke bawah serta luas daerah di kabupaten Jember, yang diperoleh dari Jember dalam Angka 2011 pada data PODES (Potensi Desa).

1. Analisis data dalam penelitian ini adalah sebagai berikut:
2. Menentukan variabel respon dan variabel penyerta dari data yang telah disajikan.
3. Mendeteksi adanya outlier pada data dengan intepretasi gambar/plot.
4. Membentuk SAE model dengan asumsi normal dan linier, yaitu : $\hat{\bar{y}}_{ij} = \hat{\theta}_{ij} = x_{ij}^T\beta + v_i + e_{ij}$ dengan *i* = *1, 2, …, m* dan asumsi v_i dan e_i saling bebas.
5. Memprediksi $\hat{\bar{y}}$ dengan EBLUP yang disebut $\hat{\bar{y}}_{EBLUP}$ untuk data dengan *outlier* dan tanpa *outlier*.
6. Memprediksi $\hat{\theta}$ dengan *Maximum Likelihood* ($\hat{\theta}_i^{REBLUP}$) dan M-*Estimation* yang ($\hat{\theta}_i^{REBLUP\,MST}$) untuk data dengan *outlier* dan tanpa *outlier*.
7. Menghitung *Mean Square Error* dari ketiga metode diatas yang selanjutnya disebut RMSE $\hat{\bar{y}}_{EBLUP}$, RMSE $\hat{\theta}_i^{REBLUP}$ dan RMSE $\hat{\theta}_i^{REBLUP\,MST}$
8. Membandingkan hasil *Root Mean Square Error* dari ketiga metode tersebut.
9. Menarik kesimpulan dari hasil tersebut.

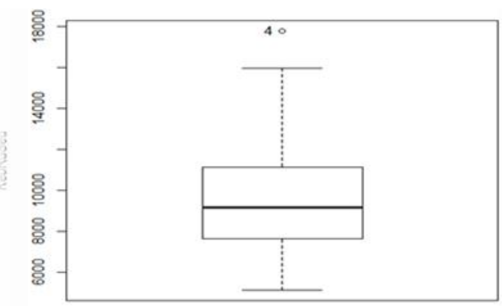
HASIL DAN PEMBAHASAN

Data yang digunakan pada penelitian ini adalah data SUSENAS 2012 dan PODES. Adapun yang menjadi variabel respon adalah jumlah kebutuhan rumah sederhana pada masing-masing kecamatan di Kabupaten Jember, kemudian variabel penyerta yang digunakan berdasarkan asumsi Real Estate Indonesia adalah jumlah rumah tangga ekonomi menengah dan luas daerah/lahan per kecamatan di Kabupaten Jember. Dari data tersebut diperoleh hasil rangkuman dalam tabel berikut ini :

Tabel 1. Hasil Rangkuman Data

Variabel		Jangkauan	Varian	Mean
Y	Kebutuhan Rumah Sederhana	12632	11712110	9761
X_1	Jumlah Keluarga Menengah ke Bawah	7878	3786476	3532
X_2	Luas Daerah	499,52	10070,86	106,24

Dari tabel 1 dapat dilihat jumlah varian yang cukup besar terutama pada variabel respon yaitu 11.712.110 . Karena pada data tersebut terdapat varian yang cukup besar maka pada data tersebut dapat kita duga adanya outlier. Selanjutnya akan dideteksi adanya outlier pada data kebutuhan rumah sederhana dengan menggunakan boxplot seperti pada gambar berikut ini :



Gambar 1. Boxplot Kebutuhan Rumah Sederhana

Dari boxplot diatas dapat dilihat bahwa terdapat outlier pada data. Outlier tersebut terdapat pada kecamatan no.4 yaitu Kecamatan Wuluhan. Setelah kita ketahui adanya outlier pada data, selanjutnya parameter pada data tersebut kita prediksi metode EBLUP dan REBLUP. Kedua metode pada pendugaan parameter ini menggunakan program R versi 3.1.3. Adapun parameter yang kita duga yaitu $\hat{\theta}$ dengan dua kondisi yaitu dengan outlier dan tanpa outlier. Pada prediksi dengan outlier ini, kecamatan Wuluhan yaitu kecamatan no.4 yang menjadi outlier dalam data diikutsertakan dalam analisis. Hasil prediksi $\hat{\theta}$ disajikan dalam tabel berikut ini :

Tabel 2. Nilai prediksi $\hat{\theta}$ dengan *outlier*

No	Nama Kecamatan	EBLUP	Robust ML	Huber M-Est
1	Kencong	8826	9125	8963
2	Gumukmas	10678	10273	10192
3	Puger	14916	13053	13141
4	Wuluhan	16877	14118	13536
5	Ambulu	13054	11684	11750
6	Tempurejo	10569	12795	12330
7	Silo	15525	14178	14269
8	Mayang	6880	7822	7689
9	Mumbulsari	8524	8867	8840
10	Jenggawah	10731	9998	10028
11	Ajung	9590	9416	9361
12	Rambipuji	9644	9403	9368
13	Balung	10030	9673	9605
14	Umbulsari	9208	9304	9201
15	Semboro	8013	8502	8344
16	Jombang	7064	7994	7794
17	Sumberbaru	13057	11967	12035
18	Tanggul	11243	11120	11076
19	Bangsalsari	14797	12919	13128
20	Panti	8728	9484	9330
21	Sukorambi	5769	7296	7025
22	Arjasa	5617	7049	6827
23	Pakusari	5943	7072	6931
24	Kalisat	9627	9163	9287
25	Ledokombo	8493	8960	9018
26	Sumberjambe	8442	9071	9004
27	Sukowono	7728	8008	8058
28	Jelbuk	7828	8419	8298
29	Kaliwates	6285	7361	7160
30	Sumbersari	57735	7158	6892
31	Patrang	13177	11348	11486

Dari tabel 2 ditunjukkan bahwa kecamatan yang membutuhkan jumlah rumah sederhana yang paling banyak menurut EBLUP adalah Kecamatan Wuluhan yaitu sebesar 16.877 unit. Sedangkan yang membutuhkan rumah sederhana paling sedikit adalah kecamatan Arjasa yaitu sebesar 5.167 unit. REBLUP ML menunjukkan bahwa Kecamatan Silo membutuhkan jumlah rumah sederhana paling banyak yaitu sebesar 14.178 unit dan Kecamatan Arjasa yang membutuhkan jumlah rumah sederhana paling sedikit yaitu sebesar 7.049 unit.

Sedangkan dengan metode REBLUP M-*Estimation* diperoleh hasil Kecamatan Silo sebagai kecamatan yang membutuhkan jumlah sederhana paling banyak yaitu sebesar 14.269 unit. Sedangkan yang membutuhkan rumah sederhana paling sedikit adalah Kecamatan Arjasa yaitu sebesar 6.827 unit.

Selanjutnya dilakukan prediksi jumlah rumah sederhana tanpa outlier . Prediksi ini dilakukan membuang kecamatan no.4 yaitu kecamatan Wuluhan dari data. Hasil prediksi jumlah kebutuhan rumah di kabupaten Jember tanpa outlier adalah sebagai berikut :

Tabel 3. Nilai prediksi $\hat{\theta}$ tanpa *outlier*

No	Nama Kecamatan	EBLUP	Robust ML	Huber M-Est
1	Kencong	8759	8854	8843
2	Gumukmas	10611	10005	10074
3	Puger	14851	12787	13023
4	Ambulu	13006	11492	11665
5	Tempurejo	10459	12351	12134
6	Silo	15493	14045	14211
7	Mayang	6867	7769	7666
8	Mumbulsari	8530	8892	8851
9	Jenggawah	10703	9884	9978
10	Ajung	9551	9257	9291
11	Rambipuji	9611	9271	9310
12	Balung	9974	9444	9505
13	Umbulsari	9160	9110	9116
14	Semboro	7960	8287	8249
15	Jombang	7016	7798	7708
16	Sumberbaru	13028	11847	11982
17	Tanggul	11212	10997	11022
18	Bangsalsari	14794	12906	13123
19	Panti	8693	9343	9269
20	Sukorambi	5719	7093	6935
21	Arjasa	5586	6926	6773
22	Pakusari	5937	7048	6920
23	Kalisat	9663	9312	9352
24	Ledokombo	8551	9196	9122
25	Sumberjambe	8445	9081	9008
26	Sukowono	7771	8181	8134
27	Jelbuk	7801	8307	8249
28	Kaliwates	6245	7197	7088
29	Sumbersari	5682	6941	6797
30	Patrang	13140	11196	11419

Dari tabel 3 ditunjukkan bahwa kecamatan yang membutuhkan jumlah rumah sederhana yang paling banyak menurut EBLUP adalah kecamatan Silo yaitu sebesar 15.493 unit. Sedangkan yang membutuhkan rumah sederhana paling sedikit adalah kecamatan Arjasa yaitu sebesar 5.586 unit. REBLUP ML menunjukkan bahwa kecamatan Silo membutuhkan rumah sederhana paling banyak yaitu sebesar 14.045 unit dan kecamatan Arjasa yang membutuhkan rumah sederhana paling sedikit yaitu sebesar 6.926 unit. Metode REBLUP M-Estimation menunjukkan bahwa kecamatan Silo sebagai kecamatan yang membutuhkan jumlah sederhana paling banyak yaitu sebesar 14.211 unit. Sedangkan yang membutuhkan rumah sederhana paling sedikit adalah kecamatan Arjasa yaitu sebesar 6.7737 unit.

Setelah diperoleh prediksi $\hat{\theta}$, maka ketiga metode diatas dapat dibandingkan dengan menggunakan nilai Root MSE. Root MSE merupakan salah satu tolak ukur untuk menganalisis atau mengukur kesalahan metode peramalan. Nilai RMSE disajikan dalam tabel sebagai berikut :

Tabel 4. Nilai *Root Mean Square Error*

Metode	Dengan Outlier (n=31)	Tanpa Outlier (n=30)
EBLUP	40013,69	2801,016
REBLUP ML	1537,499	1400,508
REBLUP M-Estimation	1473,753	1279,613

Dari tabel 4 ditunjukkan bahwa Nilai RMSE untuk metode REBLUP M-Estimation lebih baik dibandingkan dengan metode EBLUP maupun REBLUP ML yaitu 1473,7 (dengan outlier) dan 1279,6 (tanpa outlier) . Selain itu, hal ini juga menunjukkan bahwa hasil untuk pendugaan yang menggunakan EBLUP ternyata mengalami lonjakan nilai RMSE yang cukup besar. RMSE awal yang diperoleh tanpa outlier yaitu 2.801 dan RMSE yang diperoleh dengan outlier yaitu 40.013. Lonjakan RMSE pada EBLUP mencapai lima kali lipat dengan satu outlier saja yang diikut sertakan dalam analisis data. Sedangkan untuk REBLUP M-Estimation relatif lebih stabil hasil RMSEnya. Sehingga dapat disimpulkan bahwa REBLUP M-Estimation lebih robust terhadap outlier.

KESIMPULAN

Dari hasil diatas dapat disimpulkan bahwa REBLUP *M-Estimation* lebih robust terhadap outlier karena hasil untuk pendugaan yang menggunakan EBLUP ternyata mengalami lonjakan nilai RMSE yang cukup besar, yaitu dari 2.801 menjadi 40.013 dengan satu outlier saja yang diikut sertakan dalam analisis data. Sedangkan untuk REBLUP *M-Estimation* relatif lebih stabil hasil RMSEnya. Untuk hasil prediksi jmlah rumah sederhana dengan menggunakan REBLUP M-Estimation diperoleh Kecamatan Silo sebagai kecamatan yang membutuhkan jumlah sederhana paling banyak yaitu sebesar 14.269 unit. Sedangkan yang membutuhkan rumah sederhana paling sedikit adalah Kecamatan Arjasa yaitu sebesar 6.827 unit.

DAFTAR PUSTAKA

Andrews, D.F. 1972. *Robust Estimate of Location Survey and Advances*. Amerika Serikat: Pricenton University

Aunuddin. 1989. *Analisis Data*. Bogor: Penerbit IPB. Bogor

Chen, C. 2002. *Robust Regressions and Outlier Detection*. New York : John Wiley and Sons.

Darsyah, M.Y, Maulana, U, dan Utami, T.W. 2013. *Small Area Estimation untuk Pendugaan Jumlah Penduduk Miskin di Kota Semarang dengan Pendekatan Kernel-Bootstrap*. Fakultas Matematika dan Ilmu Pengetahuan Alam Program Studi Statistika Universitas Muhammadiyah Semarang. Jurnal Statistika. Vol 2.

Datta, G.S dan Gosh, M.1991. *Bayesian Prediction in Linear models Application to Small Area Estimation*. The Annals of Statistics .Vol 19:1748-1170.

Dewi, L.A. 2013. *Estimasi parametermodel regresi M-Kuantil menggunakan Iterative Reweighted Least Square (IRLS)*.Tidak Diterbitkan. Skripsi. Solo : Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Solo

Fellner, W. 1986. *Robust estimation of variance components*. Technometrics. Vol 28 : 51-60.

Huber, P.J. 1981. *Robust Statistics*. New York : John Willey and Sons.

Kurnia. 2006. *Modifikasi General Regression dan Pendekatan Non Parametrik pada Pendugaan Area Kecil*. Bogor : Institut Pertanian Bogor.

Mustikasari, Erna. 2012. *Estimasi Parameter Regresi Robust dengan Metode Estimasi MM pada Produksi Cabai di Indonesia*.Tidak Diterbitkan. Skripsi. Solo : Fakultas Matematika dan Ilmu Pengetahuan Universitas Sebelas Maret.

Myers, R.H. 1990. *Classical and Modern Regression with Applications 2nd Edition*. Boston: PWS-KENT Publishing Company.

Moktarian, P dan Chambers, R. 2013. *An Outlier Robust Block Bootstrap for Small Area Estimation*. Australia : Faculty of Engineering and Information Sciences Paper University of Wolonglong.

Rao JNK. 2003. *Small Area Estimation*. New York: John Wiley and Sons.

Roesseuw, R. J and A.M . Leroy. 1987. *Robust Regression and Outlier Detection with the Robustreg Procedure, statistics and Data Analysis*. New Castle : SAS Institute.

Sadik. 2008. *Pendekatan Metode Pemulusan Kernel pada Small Area Estimation*. Fakultas Matematika dan Ilmu Pengetahuan Alam Jurusan Matematika Universitas Yogyakarta. Jurnal Statistika. Vol 8: 31-36.

Schoch, T. 2012. *Robust Small Area Estimation: a vignette*. Switzerland : University of Applied Sciences Northwestern.

Sinha, S. K. and Rao, J. N. K. 2009. *Robust small area estimation*. The Canadian Journal of Statistics, 37, 381-399.