

# SMALL AREA ESTIMATION UNTUK PEMETAAN ANGKA MELEK HURUF DI KABUPATEN REMBANG

Moh Yamin Darsyah<sup>1</sup>, Iswahyudi Joko Suprayitno<sup>2</sup>

<sup>1</sup> Program Studi Statistika, UNIMUS

Email: mydarsyah@unimus.ac.id

<sup>2</sup> Program Studi Pendidikan Matematika, UNIMUS

## ABSTRAK

*Small Area Estimation (SAE)* merupakan suatu teknik statistika untuk menduga parameter-parameter subpopulasi yang ukuran sampel nya kecil. Teknik pendugaan ini “*borrowing information*” memanfaatkan data dari domain besar (seperti data sensus data susenas) untuk menduga variabel yang menjadi perhatian pada domain yang lebih kecil yang selanjutnya dikenal pendugaan tidak langsung. Pendugaan langsung tidak mampu memberikan ketelitian yang cukup bila ukuran sampel dalam area kecil, sehingga statistik yang dihasilkan akan memiliki varian yang besar atau bahkan menghasilkan pendugaan yang bias. Penelitian *SAE* diaplikasikan untuk memetakan Angka Melek Huruf di Kabupaten Rembang. variabel angka partisipasi murni berperan signifikan dalam mempengaruhi Angka Melek Huruf daerah. Kecamatan Rembang merupakan kecamatan dengan Angka Melek Huruf tertinggi di Kabupaten Rembang sedangkan kecamatan yang memiliki Angka Melek Huruf terendah di Kecamatan Sedan.

**Kata kunci :** *Small Area Estimation, Nonparametrik, Angka Melek Huruf*

## PENDAHULUAN

Suatu area disebut area kecil apabila sampel yang diambil pada area tersebut tidak 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 kecil.

*Small Area Estimation (SAE)* merupakan suatu teknik statistika untuk menduga parameter-parameter subpopulasi yang ukuran sampel nya kecil. Teknik pendugaan ini memanfaatkan data dari domain besar seperti data sensus dan data susenas untuk menduga variabel yang menjadi perhatian pada domain yang lebih kecil. Pendugaan sederhana pada area kecil yang didasarkan pada penerapan model desain penarikan contoh (*design-based*) disebut sebagai pendugaan langsung (*direct estimation*) dimana pendugaan langsung tidak mampu memberikan ketelitian yang cukup bila ukuran sampel dalam area itu kecil sehingga statistik yang dihasilkan akan memiliki varian yang besar atau bahkan pendugaan tidak dapat

dilakukan karena berakibat estimasi yang bias (Rao, 2003).

Sebagai alternatif teknik pendugaan untuk meningkatkan efektivitas ukuran sampel dan menurunkan eror maka dikembangkan teknik pendugaan tak langsung (*indirect estimation*) untuk melakukan pendugaan pada area kecil dengan ketelitian yang cukup. Teknik pendugaan ini dilakukan melalui suatu model yang menghubungkan area terkait melalui penggunaan informasi tambahan atau variabel penyerta (*model based*) inilah yang selanjutnya dikenal dengan konsep *SAE*.

Terdapat dua masalah pokok dalam pendugaan area kecil. Masalah pertama adalah bagaimana menghasilkan suatu dugaan yang akurat dengan ukuran sampel kecil pada suatu domain atau area kecil. Masalah kedua yaitu bagaimana menduga *Mean Square Error (MSE)*. Solusi untuk masalah tersebut adalah dengan “meminjam informasi” baik dari dalam area, luar area, maupu luar survei.

Berbagai penelitian yang berkaitan dengan pendugaan area kecil dengan pendekatan nonparametrik antara lain Darsyah (2013)

menggunakan small area estimation dengan metode kernel-*bootstrap*, Indahwati, Sadik dan Nurmasari (2008) melakukan small area estimation dengan pendekatan pemulusan kernel, Opsmer (2005) melakukan small area estimation menggunakan penalized spline, Mukhopadhyay dan Maiti (2004) melakukan small area estimation dengan pendekatan nonparametrik.

Penelitian SAE diaplikasikan untuk mengestimasi Angka Melek Huruf di Kabupaten Rembang dengan pendekatan nonparametrik dengan menggunakan fungsi kernel- bootstrap. Kabupaten Rembang merupakan salah satu daerah tertinggal di Propinsi Jawa Tengah maka dibutuhkan kajian strategis dan mendalam untuk mengetahui informasi yang akurat sehingga pembangunan daerah tepat sasaran.

Dalam kebanyakan aplikasi pendugaan area kecil, digunakan asumsi model linier campuran dan pendugaannya sensitif terhadap asumsi ini. Jika asumsi kelinieran antara rata-rata area kecil dan variabel penyerta tidak terpenuhi, maka "meminjam kekuatan" dari area lain dengan menggunakan model linier tidak tepat. Mukhopadhyay dan Maiti (2004) menggunakan model

$$y_i = \theta_i + \epsilon_i \quad (1)$$

$$\theta_i = m(x_i) + u_i \quad (2)$$

Dimana  $i = 1, 2, \dots, m$  menyatakan banyaknya area kecil. Fungsi  $m(\cdot)$  adalah fungsi mulus (smoothing function) yang mendefinisikan relasi antara  $x$  dan  $y$ .  $\theta_i$  adalah rata-rata area kecil yang tidak teramati,  $y_i$  adalah pendugaan langsung dari rata-rata area kecil yang tersampel,  $u_i$  galat peubah acak yang berdistribusi independen dan identik dengan  $E(u_i) = 0$  dan  $var(u_i) = \sigma_u^2$ , dan  $\epsilon_i$  berdistribusi independen dan identik dengan  $E(\epsilon_i) = 0$  dan  $var(\epsilon_i) = D_i$ , dengan asumsi  $D_i$  diketahui. Substitusi persamaan (1) dan (2) akan menghasilkan persamaan berikut:

$$y = m(x_i) + u + \epsilon \quad (3)$$

Untuk menduga  $m(x_i)$ , Mukhopadhyay dan Maiti (2004) menggunakan pendugaan kernel Nadaraya-Watson

$$\hat{m}_h(x_i) = \frac{\sum_i K_h(x-x_i)y_i}{\sum_i K_h(x-x_i)} \quad (4)$$

Dimana  $K_h(\cdot)$  adalah fungsi kernel dengan *bandwidth*  $h$  dan  $K_h(u) = \frac{1}{h}K(u/h)$  dengan  $K_h(\cdot)$  memenuhi:

- i.  $K(\cdot)$  simetri
- ii.  $K(\cdot)$  terbatas dan kontinu pada daerah hasil  $x$
- iii.  $\int K(a)da = 1$

Fungsi kernel yang sering dipakai adalah fungsi normal dan besarnya *bandwidth* dipilih  $h \propto n^{-1/5}$  (hardle, 1994). Penduga di atas linier terhadap  $y_i$ , dan dapat ditulis sebagai

$$\hat{m}_h(x_i) = \frac{1}{m} \sum_{i=1}^m W_{hi}(x)y_i \quad (5)$$

$$\text{Dimana } W_{hi}(x) = \frac{K_h(x-x_i)}{1/m \sum_i K_h(x-x_i)}$$

Berdasarkan definisi di atas, penduga terbaik dari rata-rata area kecil  $\theta_i$  adalah

$$\hat{\theta}_i = \hat{\gamma}_i y_i + (1 - \hat{\gamma}_i) \hat{m}_h(x_i) \quad (6)$$

Dimana  $\hat{\gamma}_i = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + D_i}$  dan  $\hat{\sigma}_u^2$  merupakan penduga dari  $\sigma_u^2$ .

$$\hat{\sigma}_u^2 = \max \left\{ 0, \frac{1}{m-1} \sum_{i=1}^m W_{hi}(x) [y_i - \hat{m}_h(x_i)]^2 - D \right\} \quad (7)$$

Untuk *MSE* pendugaan area kecil berikut:

$$MSE(\hat{\theta}_i) = \frac{D\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + D} + (1 - \gamma)^2 mse[\hat{m}_h(x_i)] + 2D^2(\hat{\sigma}_u^2 + D)^{-3} mse(\hat{\sigma}_u^2) \quad (8)$$

Namun demikian pendugaan *MSE* diatas mempunyai kelemahan karena ada informasi yang terputus dan tidak ada rumus jadinya, maka untuk pendugaan *MSE* bisa dilakukan dengan pendekatan bootstrap berikut persamaannya

$$mse^*(\hat{\theta}_i) = \frac{1}{J} \sum_{j=1}^J (\hat{\theta}_i^{*(j)} - \theta_i^{*(j)})^2 \quad (9)$$

dimana  $J$  adalah banyaknya populasi bootstrap,  $\hat{\theta}_i^{*(j)}$  adalah penduga rata-rata area kecil  $ke-i$  dari populasi bootstrap  $ke-j$ , dan  $\theta_i^{*(j)}$  adalah nilai

sebenarnya rata-rata area kecil  $ke-i$  dari populasi bootstrap  $ke-j$ .

## METODE

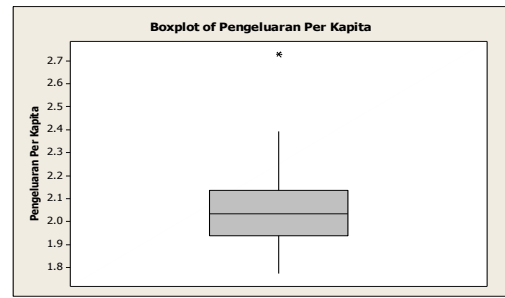
Selanjutnya akan dijelaskan mengenai Kajian Model *SAE*, Sumber Data dan Variabel Penelitian, Metodologi Penelitian sebagai berikut.

Model area kecil biasanya menggunakan model linier campuran dalam bentuk  $y = X\beta + Zu + e$ , dimana  $X$  adalah matriks peubah penyerta,  $Z$  adalah vektor acak yang biasa dikenal sebagai pengaruh area kecil, dan  $e$  adalah vektor dari galat sampel (Rao, 2003). Penduga rata-rata area kecil di duga dengan melalui estimator *best linier unbiased predictor (BLUP)*  $\hat{\theta}_i = \gamma_i \hat{\theta}_i + (1 - \gamma_i) z_i^T \tilde{\beta}$ . Mukhopadhyay dan Maiti (2004) menggunakan pendugaan kernel Nadaraya-Watson  $\hat{m}_h(x_i) = \frac{\sum_i K_h(x-x_i)y_i}{\sum_i K_h(x-x_i)}$ , dimana estimator pendugaan untuk area kecil diberikan  $\hat{\theta}_i = \hat{\gamma}_i y_i + (1 - \hat{\gamma}_i) \hat{m}_h(x_i)$ , dengan  $\hat{\gamma}_i = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + D_i}$ . Dimana nilai  $\hat{\gamma}_i$  merupakan estimator pembobot yang didapat dari *best linier unbiased predictor (BLUP)* yang merupakan bentuk parametrik. Jadi ada kemungkinan untuk penduga area kecil diatas merupakan penduga semiparametrik karena terdapat 2 (dua) pendekatan yaitu pendekatan nonparametrik dan pendekatan parametrik.

Sumber data yang digunakan dalam penelitian ini untuk Angka Melek Huruf berasal dari data Survei Sosial Ekonomi Nasional (SUSENAS) BPS Tahun 2015 dan Rembang Dalam Angka 2014. Dimana variabel respon yang menjadi perhatian dalam penelitian ini adalah Angka Melek Huruf pada masing-masing kecamatan di Kabupaten Rembang. Variabel penyerta yang digunakan yaitu Jumlah pengeluaran per kapita, Rasio tenaga pendidik, Angka partisipasi murni.

## HASIL PENELITIAN

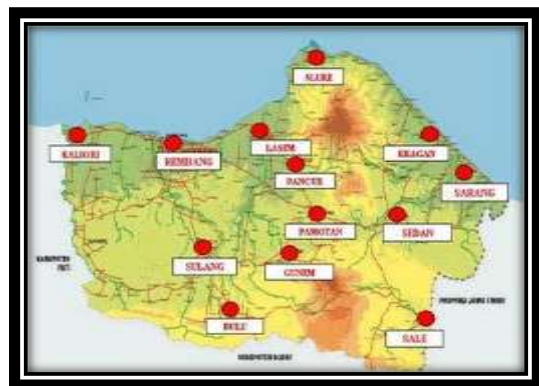
### a. Sebaran Data



Gambar 1. *Boxplot* Pendugaan Angka Melek Huruf dengan *SAE*

Gambar 1 menunjukkan adanya pencilon pendugaan yaitu Kecamatan Kota Rembang. Lebar bagian atas dan bagian bawah sama hal ini mengindikasikan bahwa keragaman Angka Melek Huruf pada kecamatan di Kabupaten Rembang tidak terlalu mencolok.

### b. Uji Keragaman Spasial (*Heterokedastisitas*)



Gambar 2. Peta Kabupaten Rembang Pengujian keragaman spasial menggunakan uji *Breusch-Pagan (BP)* menghasilkan nilai BP sebesar 5,16 dengan nilai-p (0,041) yang kurang dari taraf nyata 5%, sehingga diperoleh keputusan tolak  $H_0$  yang berarti bahwa terdapat keragaman spasial pada data Angka Melek Huruf pada tiap kecamatan di Kabupaten Rembang. Adanya keragaman spasial pada Angka Melek Huruf tersebut menunjukkan bahwa setiap kecamatan di Kabupaten Rembang memiliki karakteristik tersendiri, sehingga diperlukan pendekatan lokal untuk memodelkan dan untuk mengatasi keragaman yang terjadi pada Angka Melek Huruf.

### c. Koefisien Determinasi

Nilai  $R^2$  yang diperoleh dari model sebesar 80,49%. Hal ini berarti keragaman Angka Melek Huruf disebabkan oleh pengeluaran per kapita, rasio tenaga pendidik, dan angka partisipasi murni sebesar 80,49%, sedangkan 19,51 % sisanya disebabkan oleh adanya faktor lainnya yang turut mempengaruhi besarnya Angka Melek Huruf.

### d. Pengujian Model

*Goodness of fit* atau pengujian kesesuaian untuk model dilakukan untuk mengetahui faktor lokasi yang berpengaruh terhadap tingkat Angka Melek Huruf di Kabupaten Rembang.

Tabel 2. Uji Kesesuaian Model

	SSE	Df	F <sub>hitung</sub>	P <sub>value</sub>
Model	43,977	9,023	2,775	0,004

Berdasarkan tabel 2 di atas didapatkan nilai  $p_{value}$  (0,004) yang berarti nilai  $p_{value}$  kurang dari taraf nyata 5% ( $0,004 < 0,05$ ). Hal ini berarti tolak  $H_0$  karena nilai  $p_{value}$  lebih kecil dari taraf nyata 5%, yang artinya ada faktor pengaruh geografis pada model.

Tabel 3. Signifikansi Parameter

No	Kecamatan	Variabel
1	Bulu	$X_{2,3}$
2	Gunem	-
3	Kaliori	$X_{1,2,3}$
4	Kragan	$X_{1,2,3}$
5	Lasem	$X_{1,2,3}$
6	Pamotan	$X_{2,3}$
7	Pancur	$X_{2,3}$
8	Rembang	$X_{1,2,3}$
9	Sale	$X_3$
10	Sarang	$X_{1,3}$
11	Sedan	-
12	Sluke	-
13	Sulang	$X_{2,3}$
14	Sumber	-

Berdasarkan tabel 3 diperoleh hasil bahwa ada 10 kecamatan yang dipengaruhi oleh variabel angka partisipasi murni ( $X_3$ ) yang dominan mempengaruhi angka melek huruf di Kabupaten Rembang. Ada 4 kecamatan yang tidak di pengaruhi oleh ketiga variabel yang

digunakan dalam penelitian ini. Penelitian selanjutnya perlu mengkaji variable variabel yang lebih komprehensif yang ini diduga secara signifikan yang mempengaruhi Angka Melek Huruf di Kabupaten Rembang.

### KESIMPULAN

Berikut kesimpulan yang diperoleh dari hasil penelitian yang telah dilakukan antara lain:

- Aplikasi Small Area Estimation dapat di kombain dengan Regresi Spasial.
- Variabel yang paling berpengaruh terhadap Angka Melek Huruf di Kabupaten Rembang secara keseluruhan yaitu angka partisipasi murni walaupun variabel rasio tenaga pendidik cukup besar.
- faktor lokasi/area antar kecamatan berpengaruh terhadap tingkat Angka Melek Huruf di Kabupaten Rembang.
- Nilai  $R^2$  yang diperoleh dari model sebesar 80,49%. Hal ini berarti keragaman angka melek huruf disebabkan oleh besarnya pengeluaran per kapita, rasio tenaga pendidik, dan angka partisipasi murni sebesar 80,49%, sedangkan 19,51 % sisanya disebabkan oleh adanya faktor lainnya yang turut mempengaruhi Angka Melek Huruf.

### DAFTAR PUSTAKA

- [BPS]. Badan Pusat Statistik. (2015). <http://www.bps.go.id/glossary/2015>.
- Darsyah, M.Y, Rumiati, A.T, Otok, B.W. (2012). *Small Area Estimation terhadap Pengeluaran Per Kapita di Kabupaten Sumenep dengan pendekatan Kernel-Bootstrap*. Prosiding Seminar Nasional MIPA UNESA, Surabaya.
- Darsyah, M.Y dan Wasono, R. (2013). *Pendugaan Tingkat Kemiskinan di Kabupaten Sumenep dengan pendekatan SAE*. Prosiding Seminar Nasional Statistika UII, Yogyakarta.
- Darsyah, M.Y dan Wasono, R. (2013). *Pendugaan IPM pada Area Kecil di Kota Semarang dengan Pendekatan Nonparametrik*. Prosiding Seminar Nasional Statistika UNDIP. Semarang.
- Darsyah, M.Y. (2013). *Small Area Estimation terhadap Pengeluaran Per Kapita di*

*Kabupaten Sumenep dengan Pendekatan Nonparametrik. Jurnal Statistika. Vol 1 Nomer 2.*

Hardle, W. (1994). *Applied Nonparametric Regression*. New York: Cambridge University Press.

Indahwati dkk (2008). *Pendugaan Area Kecil Pengeluaran Per Kapita dengan Pemulusan Kernel*. Prosiding Seminar Nasional MIPA UNY. Yogyakarta.

Mukhopadhyay P, Maiti T. (2004). *Two Stage Non-Parametric Approach for Small Area Estimation*. Proceedings of ASA Section on Survey Research Methods: 4058-4065.

Opsomer et al. (2004). *Nonparametric Small Area Estimation Using Penalized Spline Regression*. Proceedings of ASA Section on Survey Research Methods: 1-8.

Rao JNK. (2003). *Small Area Estimation*. New Jersey: John Wiley & Sons, Inc.