

PENDUGAAN BERBASIS MODEL UNTUK KASUS BINER PADA SMALL AREA ESTIMATION

Kismiantini

Jurusan Pendidikan Matematika, Universitas Negeri Yogyakarta
Karangmalang, Yogyakarta 55281, Indonesia

e-mail : kismi_uny@yahoo.com

ABSTRAK

Small Area Estimation (SAE) berguna untuk menduga parameter subpopulasi (area) yang berukuran sampel kecil. Parameter subpopulasi dalam *SAE* ini dapat berasal dari peubah respons kontinu atau peubah respons diskret (seperti biner, cacahan). Berbagai metode telah dikembangkan untuk melakukan pendugaan terhadap parameter area tersebut khususnya metode berbasis model. Pada makalah ini, akan dibahas pendugaan komposit berbasis model untuk kasus biner yaitu pendugaan berbasis model linear, berbasis model logit dan berbasis model logit dengan pengaruh acak area dapat digunakan. Pendugaan berbasis model ini mampu memberikan hasil pendugaan pada area yang apabila dengan pendugaan langsung diperoleh penduga sebesar nol.

Kata-kata kunci : *small area estimation*, kasus biner, metode berbasis model.

1. Pendahuluan

Small area estimation (SAE) merupakan suatu teknik statistika untuk menduga parameter-parameter subpopulasi yang ukuran sampelnya kecil (Rao, 2003). Di Indonesia, subpopulasi tersebut dapat berupa provinsi, kabupaten/kota, kecamatan atau kelurahan/desa. Secara umum ada tiga pendekatan untuk mendapatkan penduga parameter dalam *SAE* yaitu pendugaan langsung (*direct estimation*), pendugaan tak langsung (*indirect estimation*) dan pendugaan komposit (*composite estimation*). Bila ukuran contoh pada subpopulasi kecil bahkan nol maka statistik dari pendugaan langsung akan memiliki ragam galat yang besar bahkan pendugaan tidak dapat dilakukan (Rao, 1999). Untuk mengatasi hal tersebut, dapat digunakan pendugaan tak langsung (penduga sintetik). Sedangkan pendugaan komposit adalah pendugaan yang dilakukan dengan memboboti antara penduga langsung dengan penduga tak langsung. Pendugaan ini untuk menyeimbangkan bias dari penduga tak langsung sintetik dengan ketidakstabilan dari penduga langsung yaitu dengan memberikan rata-rata terboboti untuk kedua penduga tersebut (Ghosh & Rao, 1994).

Dalam SAE, peubah respons dapat dikategorikan dalam dua jenis yaitu peubah kontinu dan peubah diskret. Bila ingin melakukan pendugaan tentang pendapatan per kapita di suatu wilayah maka peubah respons yang menjadi perhatian merupakan peubah kontinu. Sedangkan bila ingin mengetahui proporsi pengangguran di suatu wilayah, banyaknya orang yang menderita demam berdarah maka peubah respons yang menjadi perhatian merupakan peubah diskret. Suatu peubah respons yang menyatakan “sukses” atau “gagal” sering disebut sebagai peubah biner. Jenis dari peubah respons ini akan mempengaruhi bagaimana melakukan pendugaan terhadap parameternya. Berbagai metode SAE telah dikembangkan menurut jenis dari peubah yang menjadi perhatian, khususnya menyangkut metode yang berbasis model (*model-based area estimation*) yaitu diantaranya untuk peubah biner adalah metode berbasis model linear, berbasis model logit, berbasis model logit dengan pengaruh acak area. Pada makalah ini akan dibahas tentang pendugaan proporsi status kepemilikan kartu sehat di 14 kecamatan di kota Yogyakarta dengan metode berbasis model.

2. Pendugaan Langsung

Suatu peubah respon yang menyatakan “sukses” atau “gagal” disebut sebagai peubah biner. Pada SAE untuk kasus biner, peubah yang menjadi perhatian berupa proporsi. Penduga langsung bagi proporsi merupakan penduga kemungkinan maksimum (Rao, 2003) yaitu $\hat{p}_i = y_i/n_i$, dengan mengasumsikan peubah respons diasumsikan menyebar binomial, $y_i \stackrel{iid}{\sim} \text{Binomial}(n_i, p_i)$. Penduga langsung ini mempunyai ragam yang besar karena hanya berdasarkan jumlah objek survei yang terdapat pada area kecil tersebut.

3. Pendugaan Sintetik Berbasis Model Linear

Penduga sintetik berbasis model linear untuk proporsi mengikuti model berikut (Arrieta, 2008):

$$p_i = \mathbf{x}^T \boldsymbol{\beta} + \varepsilon_i \quad (1)$$

dengan

p_i adalah proporsi sukses pada area kecil ke- i

\mathbf{x} adalah $(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$ vektor peubah penyerta yang diperoleh dari data populasi

ε_i adalah galat yang bersifat acak, $\varepsilon_i \sim N(0, \sigma^2)$.

Penduga untuk masing-masing area kecil ke- i diperoleh berikut

$$\hat{p}_i = \mathbf{x}^T \mathbf{b} \quad (2)$$

dengan $\mathbf{b} = (\mathbf{x}^T \mathbf{x})^{-1} (\mathbf{x}^T \mathbf{y})$.

4. Pendugaan Sintetik Berbasis Model Logit

Penduga sintetik berbasis model logit untuk proporsi mengikuti model berikut (Arrieta, 2008):

$$\text{logit}(p_i) = \log\left(\frac{p_i}{1-p_i}\right) = \mathbf{x}^T \boldsymbol{\beta} \quad (3)$$

dengan

p_i adalah proporsi sukses pada area kecil ke- i

\mathbf{x} adalah $(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$ vektor peubah penyerta yang diperoleh dari data populasi

Penduga untuk masing-masing area kecil ke- i diperoleh berikut

$$\hat{p}_i = \frac{\exp(\mathbf{x}^T \mathbf{b})}{1 + \exp(\mathbf{x}^T \mathbf{b})} \quad (4)$$

5. Pendugaan Sintetik Berbasis Model Logit dengan Pengaruh Acak Area

Penduga sintetik berbasis model logit dengan pengaruh acak area untuk proporsi mengikuti model berikut (Arrieta, 2008):

$$\text{logit}(p_i) = \log\left(\frac{p_i}{1-p_i}\right) = \mathbf{x}^T \boldsymbol{\beta} + u_i \quad (5)$$

dengan

p_i adalah proporsi sukses pada area kecil ke- i

u_i adalah pengaruh acak area kecil ke- i , $u_i \sim N(0, \sigma_u^2)$

\mathbf{x} adalah $(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$ vektor peubah penyerta yang diperoleh dari data populasi

Penduga *Empirical Bayes* untuk masing-masing area kecil ke- i diperoleh berikut

$$\hat{p}_i = \frac{\exp[(\mathbf{x}^T \mathbf{b}) + \hat{u}_i]}{1 + \exp[(\mathbf{x}^T \mathbf{b}) + \hat{u}_i]} \quad (6)$$

6. Pendugaan Komposit

Penduga komposit untuk proporsi diperoleh dengan memboboti penduga langsung dengan penduga tak langsung (sintetik) dengan pembobot dari Ghosh & Maiti (2004) yaitu

$$w_i = \frac{1}{1 + \lambda/n_i} \quad (7)$$

Menurut Sinha (2004), nilai λ yang sering digunakan adalah 0.5 atau 1.

Tabel 1. Penduga Komposit untuk Proporsi

Penduga Komposit	
Model Linear	$\hat{p}_i = w_i \frac{y_i}{n_i} + (1 - w_i) \mathbf{x}^T \mathbf{b}$
Model Logit	$\hat{p}_i = w_i \frac{y_i}{n_i} + (1 - w_i) \left(\frac{\exp(\mathbf{x}^T \mathbf{b})}{1 + \exp(\mathbf{x}^T \mathbf{b})} \right)$
Model Logit dengan Pengaruh Acak Area	$\hat{p}_i = w_i \frac{y_i}{n_i} + (1 - w_i) \left(\frac{\exp[(\mathbf{x}^T \mathbf{b}) + \hat{u}_i]}{1 + \exp[(\mathbf{x}^T \mathbf{b}) + \hat{u}_i]} \right)$

7. Analisis Data

Data yang diamati adalah data sekunder tentang status kepemilikan kartu sehat yang diperoleh dari Survei Sosial Ekonomi Nasional (SUSENAS) 2003 dengan materi informasi berbasis rumahtangga dan PODES 2003. Data ini diambil dari 14 kecamatan di kota Yogyakarta. Peubah yang menjadi perhatian adalah proporsi status kepemilikan kartu sehat, peubah pengamatan y_i adalah jumlah rumahtangga pemilik kartu sehat pada kecamatan ke- i , dan n_i adalah jumlah rumahtangga pada kecamatan ke- i . Peubah penyertanya adalah peubah-peubah yang diasumsikan mempengaruhi proporsi status kepemilikan kartu sehat pada suatu area, meliputi: persentase rumah tangga prasejahtera dan sejahtera 1, persentase rumah tangga pelanggan listrik PLN, dan persentase rumah tangga pelanggan telepon.

Analisis menggunakan SAS 9.1 meliputi: MS EXCEL untuk memperoleh penduga proporsi, PROC GENMOD untuk mendapatkan \mathbf{b} untuk model linear dan model logit, PROC GLIMMIX untuk mendapatkan \mathbf{b} dan \hat{u}_i untuk model logit dengan pengaruh acak area.

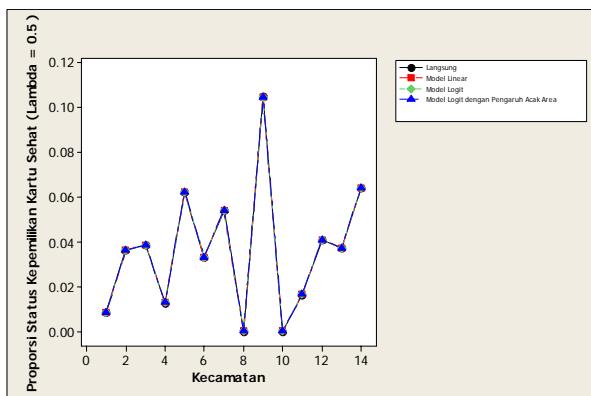
Tabel 2. Pendugaan Proporsi Status Kepemilikan Kartu Sehat ($\lambda = 0.5$)

No	Kecamatan	n_i	y_i	Penduga			
				Langsung	Model Linear	Model Logit	Model Logit dengan Pengaruh Acak Area
1	Mantrijeron	117	1	0.0085	0.0087	0.0087	0.0087
2	Kraton	110	4	0.0364	0.0364	0.0364	0.0364
3	Mergangsan	180	7	0.0389	0.0389	0.0389	0.0389
4	Umbulharjo	307	4	0.0130	0.0130	0.0130	0.0131
5	Kotagede	128	8	0.0625	0.0624	0.0624	0.0625
6	Gondokusuman	270	9	0.0333	0.0333	0.0333	0.0333
7	Danurejan	129	7	0.0543	0.0542	0.0542	0.0543
8	Pakualaman	62	0	0.0000	0.0003	0.0003	0.0004
9	Gondomanan	57	6	0.1053	0.1047	0.1047	0.1048
10	Ngampilan	41	0	0.0000	0.0005	0.0005	0.0007
11	Wirobrajan	120	2	0.0167	0.0168	0.0168	0.0168
12	Gedongtengen	97	4	0.0412	0.0412	0.0412	0.0412
13	Jetis	161	6	0.0373	0.0373	0.0373	0.0373
14	Tegalrejo	202	13	0.0644	0.0643	0.0643	0.0643

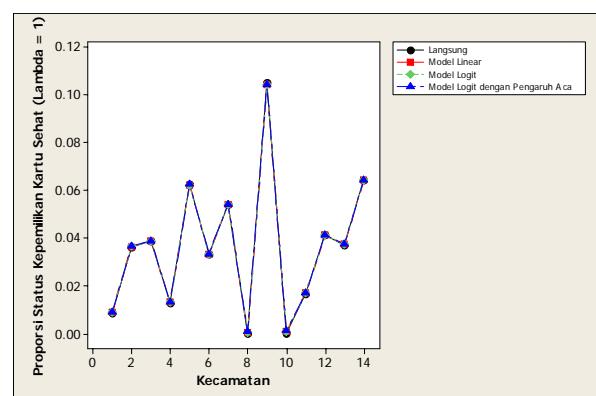
Tabel 3. Pendugaan Proporsi Status Kepemilikan Kartu Sehat ($\lambda = 1$)

No	Kecamatan	n_i	y_i	Penduga			
				Langsung	Model Linear	Model Logit	Model Logit dengan Pengaruh Acak Area
1	Mantrijeron	117	1	0.0085	0.0088	0.0088	0.0089
2	Kraton	110	4	0.0364	0.0364	0.0364	0.0365
3	Mergangsan	180	7	0.0389	0.0389	0.0389	0.0390
4	Umbulharjo	307	4	0.0130	0.0131	0.0130	0.0131
5	Kotagede	128	8	0.0625	0.0623	0.0623	0.0624
6	Gondokusuman	270	9	0.0333	0.0333	0.0333	0.0334
7	Danurejan	129	7	0.0543	0.0542	0.0542	0.0542
8	Pakualaman	62	0	0.0000	0.0007	0.0007	0.0008
9	Gondomanan	57	6	0.1053	0.1042	0.1042	0.1044
10	Ngampilan	41	0	0.0000	0.0010	0.0011	0.0013
11	Wirobrajan	120	2	0.0167	0.0169	0.0169	0.0170
12	Gedongtengen	97	4	0.0412	0.0412	0.0411	0.0412
13	Jetis	161	6	0.0373	0.0373	0.0373	0.0373
14	Tegalrejo	202	13	0.0644	0.0642	0.0642	0.0643

Secara umum dari Tabel 2 dan Tabel 3 dapat dilihat bahwa penduga proporsi status kepemilikan kartu sehat dari penduga komposit berbasis model linear, berbasis model logit dan berbasis model logit dengan pengaruh acak area relatif sama. Bahkan relatif sama dengan penduga langsung, kecuali untuk kecamatan Pakualaman dan Ngampilan. Pada kecamatan Pakualaman dan Ngampilan, dari jumlah sampel (rumahtangga) yang terpilih, jumlah rumahtangga pemilik kartu sehat tidak ada (nol) sehingga penduga langsung untuk proporsi status kepemilikan kartu sehat bernilai nol, padahal belum tentu di kecamatan tersebut semua rumahtangga tidak memiliki kartu sehat.



Gambar 1. Plot Proporsi Status Kepemilikan Kartu Sehat untuk Masing-masing Kecamatan ($\lambda=0.5$)



Gambar 2. Plot Proporsi Status Kepemilikan Kartu Sehat untuk Masing-masing Kecamatan ($\lambda=1$)

Dari Gambar 1 dan Gambar 2 dapat diketahui bahwa penduga komposit berbasis model linear, berbasis model logit dan berbasis model logit dengan pengaruh acak area relatif sama dengan penduga langsung yang ditunjukkan oleh titik-titik dan garis-garis yang berhimpit, hal ini dapat disebabkan oleh peubah-peubah penyerta yang dipilih kurang memberikan pengaruh terhadap peubah pengamatan.

8. Simpulan

Small area estimation untuk kasus biner dengan pendugaan komposit berbasis model linear, berbasis model logit dan berbasis model logit dengan pengaruh acak area mampu memberikan hasil pendugaan pada area yang apabila dengan pendugaan langsung diperoleh penduga sebesar nol. Penduga komposit berbasis model linear, berbasis model logit, berbasis model logit dengan pengaruh acak area akan memberikan hasil yang relatif sama dengan penduga langsung diantaranya disebabkan oleh pemilihan peubah penyerta yang kurang berpengaruh terhadap peubah pengamatan.

9. Daftar Pustaka

- Arrieta JI. 2008. Small-area estimation in the survey of the population in relation to activity in the A.C. of the basque country. www.eustat.es.
- Ghosh M & Rao JNK. 1994. Small area estimation: an appraisal. *Statistical Science* 9: 55-76.
- Ghosh M, Maiti T. 2004. Small-area estimation based on natural exponential family quadratic variance function models and survey weights. *Biometrika* 91: 95-112.
- Rao JNK. 1999. Some recent advances in model-based small area estimation. *Survey Methodology* 25: 175-186.
- Rao JNK. 2003. *Small Area Estimation*. New Jersey: John Wiley & Sons.
- Sinha K. 2004. Some contributions to small area estimation. Unpublished Ph.D. Dissertation. Florida: University of Florida.