PENDUGAAN AREA KECIL TERHADAP ANGKA MELEK HURUF DI KABUPATEN KUTAI KARTANEGARA DENGAN METODE *EMPIRICAL BAYES* BERBASIS MODEL BETA-BINOMIAL

Norlatifah [1], Gandhi Pawitan [2], Enny Supartini [3]

[1]Mahasiswa Program Magister Statistika Terapan Universitas Padjadjaran
[2]Staf Pengajar Universitas Katolik Parahyangan
[3]Staf Pengajar Universitas Padjadjaran
Email : [1]leti_stat@yahoo.co.id, [2]gandhi_p@unpar.ac.id , [3]arthinii@yahoo.com

Abstrak

Indeks Pembangunan Manusia (IPM) digunakan untuk mengukur keberhasilan pembangunan manusia di setiap negara. Secara nasional setiap tahunnya, Badan Pusat Statistik telah melakukan perhitungan IPM tingkat nasional, provinsi sampai tingkat kabupaten/kota. Namun sejak diterapkannya kebijakan otonomi daerah, dibutuhkan perhitungan IPM untuk wilayah yang lebih kecil seperti kecamatan. Ketidaktersediaan IPM pada tingkat kecamatan salah satunya disebabkan karena terbatasnya informasi (data) untuk perhitungan nilai komponennya pada tingkat kecamatan. Salah satu upaya untuk mengoptimalkan penggunaan data yang tersedia dan memperoleh estimasi wilayah kecil adalah dengan mengaplikasikan metode pendugaan yaitu *Small Area Estimation* (SAE). Salah satu komponen penyusun IPM yaitu angka melek huruf (AMH). Peubah kemampuan membaca dan menulis dalam perhitungan AMH merupakan peubah biner sehingga salah satu metode yang dapat diterapkan pada pendugaan area kecil adalah metode *Empirical Bayes* berbasis model Beta-Binomial. Penelitian ini bertujuan menduga angka melek huruf tingkat kecamatan di Kabupaten Kutai Kartanegara dengan metode *Empirical Bayes* berbasis model Beta-Binomial, serta akan dilihat perbandingan antara penduga langsung dan penduga *Empirical Bayes* berbasis model Beta-Binomial pada pendugaan area kecil terhadap angka melek huruf pada tingkat kecamatan di Kabupaten Kutai Kartanegara. Hasil penerapan pada angka melek huruf di kabupaten Kutai Kartanegara menunjukkan bahwa penduga *Empirical Bayes* dari model Beta-Binomial memberikan hasil pendugaan dengan ketelitian yang lebih tinggi dibandingkan penduga langsung.

Kata kunci: pendugaan area kecil, *Empirical Bayes*, Beta-Binomial

I. PENDAHULUAN

Perserikatan Bangsa-Bangsa (PBB) melalui *United Nation Development Programme* (UNDP) mengembangkan metode perhitungan *Human Development Index* (HDI) atau Indeks Pembangunan Manusia (IPM) yang digunakan untuk mengukur keberhasilan pembangunan manusia di setiap negara. IPM merupakan indeks komposit yang dihitung sebagai rata-rata sederhana dari Indeks Harapan Hidup, Indeks Pendidikan dan Indeks Standar Hidup Layak. Indeks Harapan Hidup diukur dengan angka harapan hidup pada saat bayi lahir, Indeks Pendidikan diukur dari angka melek huruf penduduk usia 15 tahun ke atas dan rata-rata lama sekolah, dan Indeks Standar Hidup Layak diukur dengan pengeluaran per kapita riil yang disesuaikan.

Secara nasional setiap tahunnya, Badan Pusat Statistik telah melakukan perhitungan IPM tingkat nasional, provinsi sampai tingkat kabupaten/kota. Namun sejak diterapkannya kebijakan otonomi daerah, dibutuhkan perhitungan IPM untuk wilayah yang lebih kecil seperti kecamatan. Hal ini diperlukan untuk membantu pemerintah daerah dalam upaya mendongkrak pembangunan di daerahnya. Ketidaktersediaan IPM pada tingkat kecamatan salah satunya disebabkan karena terbatasnya informasi (data) untuk perhitungan nilai komponennya pada tingkat kecamatan. Salah satu cara untuk mendapatkan data estimasi sampai level kecamatan adalah dengan menambah sampel yang dapat memperbaiki reliabilitas dari estimasi secara langsung pada level kecamatan. Namun demikian penambahan ukuran sampel menyebabkan biaya yang diperlukan menjadi mahal dan waktu yang diperlukan dalam survei menjadi lama. Oleh karena permasalahan tersebut, pada penelitian ini akan dilakukan penaksiran salah satu komponen penyusun IPM yaitu angka melek huruf (AMH).

Salah satu upaya untuk mengoptimalkan penggunaan data yang tersedia dan memperoleh estimasi wilayah kecil adalah dengan mengaplikasikan metode pendugaan yaitu *Small Area Estimation* (SAE). Metode SAE merupakan suatu teknik statistika untuk menduga parameter-parameter subpopulasi dengan ukuran sampel kecil. Teknik pendugaan ini memanfaatkan data dari domain besar untuk menduga parameter pada domain yang lebih kecil. Pendugaan sederhana area kecil yang didasarkan pada penerapan model desain penarikan sampel (*design-based*) disebut sebagai pendugaan langsung (*direct estimation*). Pendugaan langsung tidak mampu memberikan ketelitian yang cukup bila ukuran sampel dalam *small area* berukukan kecil, sehingga statistik yang dihasilkan akan memiliki varian yang besar atau bahkan pendugaan tidak dapat dilakukan karena tidak terwakili dalam survei (Prasad dan Rao, 1990 dalam Wahyudin, 2014).

Berbagai metode pendugaan area kecil (*small area estimation*) telah dikembangkan khususnya menyangkut metode yang berbasis model (*model-based area estimation*), antara lain *Empirical Best Linear Unbiased Predictor* (EBLUP), *Empirical Bayes* (EB), dan *Hierarchical Bayes* (HB). Metode EBLUP merupakan metode untuk data kontinu sedangkan metode EB dan HB adalah metode untuk data biner atau cacahan. Metode *Empirical Bayes* pada pendugaan area kecil pertama kali dilakukan oleh Fay dan Herriot (1979) dalam menduga pendapatan beberapa area kecil.

Peubah kemampuan membaca dan menulis dalam perhitungan AMH merupakan peubah biner sehingga salah satu metode yang dapat diterapkan pada pendugaan area kecil adalah metode *Empirical Bayes.* Metode ini bekerja dengan menggunakan inferensi dari estimasi *posterior* untuk menentukan dugaan parameter. Salah satu distribusi *posterior* yang digunakan pada metode *Empirical Bayes* yaitu distribusi Beta-Binomial. Distribusi Beta-Binomial merupakan distribusi *posterior* pada metode *Empirical Bayes* yang memiliki dua tahapan, yaitu tahap pertama diasumsikan bahwa peubah yang menjadi perhatian,

$$y_i \overset{iid}{\sim} Binomial(n_i, p_i);\; y_i = 0, \dots, n_i,\; 0 < p_i < 1,\; i = 1, \dots, m$$

sedangkan pada tahap ke dua diasumsikan bahwa $p_i \overset{iid}{\sim} Beta(\alpha, \beta)$ sebagai prior (Kismiantini, 2010).

Dalam penelitian ini akan dipelajari pemodelan SAE menggunakan pendekatan Bayesian metode *Empirical Bayes* berbasis model Beta-Binomial. Untuk studi kasus, akan diestimasi Angka Melek Huruf pada area kecil (kecamatan) di Kabupaten Kutai Kartanegara berdasarkan data Susenas 2011.

II. METODOLOGI

2.1 Penduga Langsung Berbasis Peubah Respon Binomial

Peubah respon y_{ij} merupakan peubah respon biner yang diukur pada area ke-i dimana y_{ij} bernilai 1 atau 0. Sebagai contoh y_{ij} adalah peubah yang mengukur kemampuan baca tulis, maka y_{ij} = 1 jika individu tertentu di area ke-i bisa baca dan tulis dan y_{ij} = 0 jika tidak bisa baca tulis. Jika peubah y_{ij} diasumsikan memiliki sebaran Bernoulli dengan parameter p_i, maka fungsi kepekatan peluang dari y_{ij} adalah:

$$f(y_{ij} | p_i) = p_i^{\,y_{ij}} (1 - p_i) \tag{1}$$

atau ditulis $y_{ij} | p_i \overset{ind}{\sim} Bernoulli(p_i)$, untuk j=1,2,…,n_i ; i=1,2,…,m. Selanjutnya didefinisikan $y_i = \sum_j y_{ij}$, adalah jumlah kejadian yang menjadi perhatian di area ke-i, maka y_i memiliki sebaran Binomial dengan fungsi peluang:

$$f(y_i | p_i) = \binom{n_i}{y_i} p_i^{\,y_i} (1 - p_i)^{n_i - y_i} \tag{2}$$

dengan $y_i = 0, \dots, n_i,\; 0 < p_i < 1,\; i = 1, \dots, m$

atau ditulis:

$$y_i | p_i \overset{ind}{\sim} Binomial(n_i, p_i)$$

Dalam contoh kasus penelitian ini, y_i adalah jumlah individu di area ke-i yang bisa membaca dan menulis.

Parameter area kecil yang ingin diduga adalah proporsi area kecil: $p_i = \bar{Y}_i = \sum_j \frac{y_{ij}}{N_i}$

Jika penarikan sampel dilakukan dengan metode acak sederhana, maka penduga proporsi di area ke-*i* yaitu $\hat{p}_i$, diturunkan melalui metode pendugaan peluang maksimum/ *maximum likelihood* (ML), yaitu $\hat{p}_i = \sum_j \frac{y_{ij}}{n_i} = \frac{y_i}{n_i}$. Penduga ML ini merupakan pendugaan langsung melalui pendekatan klasik (Rumiati, 2012).

Penduga ini merupakan penduga kemungkinan maksimum yang bersifat tak-bias karena nilai harapan dari penduga sama dengan parameternya.

$$E(\hat{p}_i) = E\left(\frac{y_i}{n_i}\right) = \frac{1}{n_i}E(y_i) = \frac{1}{n_i}n_i p_i = p_i$$

Sehingga dugaan kuadrat tengah galat atau *mean square error* (MSE) sama dengan ragamnya, yaitu

$$MSE(\hat{p}_i) = \hat{V}ar(\hat{p}_i) \qquad \text{(Kismiantini, 2010)}$$

Melalui pendekatan Bayes, pendugaan parameter p_i dapat dilakukan secara langsung yaitu dengan tidak memanfaatkan informasi tambahan dari peubah penyerta dan pendugaan tidak langsung yaitu menggunakan model dengan memanfaatkan informasi dari peubah penyerta.

Pendugaan langsung melalui pendekatan Bayes adalah menganggap parameter p_i merupakan peubah yang memiliki distribusi tertentu. Dalam pendugaan Bayes terdapat dua jenis informasi yaitu informasi *prior* diperoleh dari sebaran *prior* dan informasi dari hasil survei. Untuk peubah Binomial, sebaran *prior* yang digunakan adalah sebaran Beta atau Logit Normal (Rumiati, 2012).

2.2 Metode *Empirical Bayes* Berbasis Model Beta-Binomial

Model ini merupakan suatu model yang berawal dari model Bernoulli dengan model peluang untuk data biner yang dinyatakan dengan y_{ij} untuk j=1,…,n_i ; i=1,2,…,m

dengan model dasar:

$$y_{ij}|p_i \overset{iid}{\sim} Bernoulli(p_i)$$

atau

$$y_i|p_i \overset{iid}{\sim} Binomial(n_i, p_i)$$

dan

$$p_i \overset{iid}{\sim} Beta(\propto, \beta);\ \ \alpha > 0,\ \ \beta > 0$$

Dengan *Beta*(α,β) menyatakan sebaran Beta dengan parameter α dan β serta fungsi kepekatan untuk p_i adalah

$$f(p_i|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p_i^{\alpha-1}(1-p_i)^{\beta-1}\ ;\ \alpha > 0, \beta > 0$$

dan $\Gamma(.)$ adalah fungsi Gamma.

Untuk menyederhanakan $\boldsymbol{y_i} = (\boldsymbol{y_{i1}}, \dots, \boldsymbol{y_{in_i}})^{\boldsymbol{T}}$ menjadi total sampel $y_i = \sum_j y_{ij}$. Diketahui bahwa $y_i|p_i \overset{iid}{\sim} Binomial(n_i, p_i)$ yang mempunyai fungsi kepekatan sebagai berikut:

$$f(y_i|p_i) = \binom{n_i}{y_i}\ {p_i}^{y_i}(1-p_i)^{n_i-y_i}$$

Berdasarkan persamaan fungsi kepekatan p_i dan fungsi kepekatan y_i maka

$$p_i|y_i, \alpha, \beta \overset{ind}{\sim} Beta(y_i + \alpha, n_i - y_i + \beta)$$

Oleh karena itu, penduga Bayes bagi p_i adalah

$$\hat{p}_i^B(\alpha,\beta) = E(p_i|y_i,\alpha,\beta) = \frac{y_i+\alpha}{n_i+\alpha+\beta}$$

dan ragam *posterior* bagi p_i adalah:

$$V(p_i|y_i,\alpha,\beta) = \frac{(y_i+\alpha)(n_i-y_i+\beta)}{(n_i+\alpha+\beta+1)(n_i+\alpha+\beta)^2}$$

Sebaran penghubung $f(p_i|\alpha,\beta)$ dinamakan *prior konjugate* pada sebaran *posterior*, $f(p_i|y_i,\alpha,\beta)$ mempunyai bentuk yang sama dengan sebaran *prior*nya. Dari sebaran penghubung tersebut, maka digunakan model yang sering disebut model Beta-Binomial dengan sebaran peluang marginal:

$$f(y_i|n_i,\alpha,\beta) = \binom{n_i}{y_i}\frac{\Gamma(\alpha+y_i)\Gamma(\beta+n_i-y_i)}{\Gamma(\alpha+\beta+n_i)}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$$

$$= \binom{n_i}{y_i}\frac{B(\alpha+y_i,\beta+n_i-y_i)}{B(\alpha,\beta)}$$

Untuk menduga parameter α dan β digunakan metode momen Kleinman:

$$\frac{\hat{\alpha}}{\hat{\alpha}+\hat{\beta}} = \hat{p}$$

dan

$$\frac{1}{\hat{\alpha}+\hat{\beta}+1} = \frac{n_T S_p^2 - \hat{p}(1-\hat{p})(m-1)}{\hat{p}(1-\hat{p})\left[n_T - \sum_i \left(\frac{n_i^2}{n_T}\right) - (m-1)\right]}$$

dengan rataan sampel berbobot:

$$\hat{p} = \sum\nolimits_i \left(\frac{n_i}{n_T}\right)\hat{p}_i$$

dan ragam sampel terboboti:

$$s_p^2 = \sum\nolimits_i \left(\frac{n_i}{n_T}\right)(\hat{p}_i-\hat{p})^2$$

dimana:

m = banyaknya area

n_i = banyaknya elemen/ individu di area ke-*i*

$n_T = \sum_i n_i$ = total elemen/ individu untuk *m* area

$\hat{p}$ = rataan sampel berbobot

s_p^2 = ragam sampel terboboti

Dugaan parameter $\hat{\alpha}$ dan $\hat{\beta}$ dinyatakan dengan rumus berikut:

$$\hat{\alpha} = \hat{p}\left[\frac{\hat{p}(1-\hat{p})\left[n_T - \sum_i\left(\frac{n_i^2}{n_T}\right) - (m-1)\right]}{n_T S_p^2 - \hat{p}(1-\hat{p})(m-1)} - 1\right] \qquad \text{dan}$$

$$\hat{\beta} = \hat{p}\left[\frac{\hat{p}(1-\hat{p})\left[n_T - \sum_i \left(\frac{n_i^2}{n_T}\right) - (m-1)\right]}{n_T S_p^2 - \hat{p}(1-\hat{p})(m-1)} - 1\right]\left[\frac{1}{\hat{p}} - 1\right]$$

dimana $\hat{\alpha}$ dan $\hat{\beta}$ dugaan parameter sebaran Beta-Binomial.

Pensubstitusian parameter $\hat{\alpha}$ dan $\hat{\beta}$ dari metode momen Kleinman ke penduga *Empirical Bayes* (EB) bagi $\hat{p}_i^{EB}$ diperoleh

$$\hat{p}_i^{EB} = \hat{p}_i^{B}\left(\hat{\alpha}, \hat{\beta}\right) = \hat{\gamma}_i \hat{p}_i + (1 - \hat{\gamma}_i)\hat{p}$$

dengan $\hat{\gamma}_i = \frac{n_i}{n_i + \hat{\alpha} + \hat{\beta}}$ dan $\hat{p}_i^{EB}$ merupakan rataan berbobot dari penduga langsung $\hat{p}_i$ dan penduga sintetik $\hat{p}$ (Rao, 2003a).

2.3 Pendugaan MSE dengan Metode Jackknife

Menurut Rao (2003a), penentuan penduga ragam (*mean square error*/ kuadrat tengah galat) dengan metode Jackknife untuk penduga EB adalah

$$MSE_J\left(\hat{\theta}_i^{EB}\right) = E(\hat{\theta}_i^{EB} - \hat{\theta}_i^{B})^2 + E(\hat{\theta}_i^{B} - \theta_i)^2 = \widehat{M}_{1i} + \widehat{M}_{2i}$$

dengan

$$\widehat{M}_{1i} = g_{1i}\left(\hat{\alpha}, \hat{\beta}, y_i\right) - \frac{m-1}{m}\sum_{l=1}^{m}\left[g_{1i}\left(\hat{\alpha}_{-l}, \hat{\beta}_{-l}, y_i\right) - g_{1i}(\hat{\alpha}, \hat{\beta}, y_i)\right]$$

$$\widehat{M}_{2i} = \frac{m-1}{m}\sum_{l=1}^{m}\left[\hat{\theta}_{i,-l}^{EB} - \hat{\theta}_i^{EB}\right]^2$$

dimana:

$\hat{\theta}_i^{EB} = k_i(\hat{\theta}_i, \hat{\alpha}, \hat{\beta})$ merupakan penduga Bayes bagi θ_i

$\hat{\theta}_{i,-l}^{EB} = k_i(\hat{\theta}_{i,-l}, \hat{\alpha}_{-l}, \hat{\beta}_{-l})$ merupakan penduga Bayes bagi $\theta_{i,-l}$

$\hat{\alpha}_{-l}$ dan $\hat{\beta}_{-l}$ adalah penduga untuk area kecil ke-*l* yang dihapus

III. HASIL DAN PEMBAHASAN

Data yang akan digunakan dalam penelitian ini adalah data sekunder yang bersumber dari Badan Pusat Statistik (BPS). Angka melek huruf sebagai peubah respon bersumber dari *raw data* hasil Survei Sosial Ekonomi Nasional (Susenas) 2011. Peubah yang menjadi perhatian adalah proporsi kemampuan baca tulis, peubah pengamatan y_i adalah jumlah penduduk usia 15 tahun ke atas yang bisa membaca dan menulis pada kecamatan ke-i, dan n_i adalah jumlah penduduk usia 15 tahun ke atas pada kecamatan ke-*i*.

Tabel 1. Pendugaan Angka Melek Huruf

No.	Nama Kecamatan	n_i	y_i	Langsung		EB Beta-Binomial	
				Penduga	MSE	Penduga	MSE
1	Semboja	178	169	0.9494	0.0003	0.9523	0.0000
2	Muara Jawa	141	136	0.9645	0.0002	0.9653	0.0000
3	Sanga Sanga	52	51	0.9808	0.0004	0.9764	0.0002
4	Loa Janan	137	136	0.9927	0.0001	0.9884	0.0001
5	Loa Kulu	138	136	0.9855	0.0001	0.9825	0.0000
6	Muara Muntai	58	52	0.8966	0.0016	0.9215	0.0012
7	Muara Wis	30	30	1.0000	0.0000	0.9843	0.0003
8	Kota Bangun	53	49	0.9245	0.0013	0.9408	0.0007
9	Tenggarong	550	541	0.9836	0.0000	0.9829	0.0001
10	Sebulu	54	54	1.0000	0.0000	0.9888	0.0002
11	Tenggarong Seberang	93	88	0.9462	0.0005	0.9519	0.0003
12	Anggana	63	57	0.9048	0.0014	0.9257	0.0010
13	Muara Badak	44	44	1.0000	0.0000	0.9873	0.0002
14	Marang Kayu	23	23	1.0000	0.0000	0.9823	0.0003
15	Muara Kaman	112	110	0.9821	0.0002	0.9793	0.0000
16	Kenohan	33	27	0.8182	0.0045	0.8907	0.0029
17	Kembang Janggut	18	18	1.0000	0.0000	0.9805	0.0004
18	Tabang	26	26	1.0000	0.0000	0.9832	0.0003

Dari Tabel 1 terlihat bahwa nilai penduga langsung dan penduga *Empirical Bayes* berbasis model Beta-Binomial memberikan nilai dugaan yang tidak jauh berbeda, bahkan relatif sama, kecuali untuk kecamatan Muara Wis, Sebulu, Muara Badak, Marang Kayu, Kembang Janggut dan Tabang. Pada kecamatan-kecamatan tersebut, semua penduduk usia 15 tahun ke atas yang menjadi sampel adalah mampu membaca dan menulis sehingga penduga langsung untuk proporsi melek huruf-nya bernilai 1 (satu), padahal belum tentu di kecamatan tersebut semua penduduk usia 15 tahun ke atas-nya mampu membaca dan menulis.

Dilihat dari nilai MSE, secara umum penduga langsung memberikan *mean square error* (MSE) yang lebih besar sehingga mempunyai ketelitian yang rendah, hal ini disebabkan jumlah sampel yang relatif kecil. Sedangkan penduga *Empirical Bayes* berbasis model Beta-Binomial memberikan hasil pendugaan dengan ketelitian yang lebih tinggi yang ditunjukkan oleh semakin kecilnya nilai MSE.

IV. KESIMPULAN

Pendugaan area kecil dengan metode *Empirical Bayes* berbasis model Beta-Binomial menghasilkan kualitas penduga yang lebih baik dibanding penduga langsung yang ditunjukkan dengan kecilnya nilai MSE hasil pendugaan.

DAFTAR PUSTAKA

BPS. 2008. *Indeks Pembangunan Manusia 2006-2007*. Jakarta.

Fay, R.E. & Herriot, R.A. 1979. Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data. *Journal of the American Statistical Association,* 74: 269-277.

Ghosh, M. & Rao, J.N.K. 1994. Small Area Estimation: An appraisal. *Statistical Science,* 9: 55–93.

Kismiantini. 2010. Penerapan Metode Bayes Empirik pada Pendugaan Area Kecil untuk Kasus Biner (Studi tentang Proporsi Status Kepemilikan Kartu Sehat di Kota Yogyakarta). *In: Seminar Nasional Penelitian, Pendidikan dan Penerapan MIPA Universitas Negeri Yogyakarta*.

Rao, J.N.K. 2003a. *Small Area Estimation.* New Jersey: John Wiley & Sons.

Rao, J.N.K. 2003b. Some New Developments in Small Area Estimation. *JIRSS,* 2: 145-169.

Rumiati, A.T. 2012. Model Bayes Untuk Pendugaan Area Kecil Dengan Penarikan Contoh Berpeluang Tidak Sama Pada Kasus Respon Binomial dan Multinomial. *Disertasi*. Institut Pertanian Bogor.

Wahyudin, Y. 2014. Small Area Estimation Pendekatan Empirical Bayes Berbasis Model Poisson-Gamma (Studi Kasus Data Dasar Penghitungan Angka Harapan Hidup di Kabupaten Sumbawa). *Tesis*. Institut Teknologi Sepuluh Nopember.