

# Pemodelan Angka Partisipasi Kasar Perguruan Tinggi Menggunakan *Small Area Estimation* dengan Pendekatan Semiparametrik *Penalized Spline* di Provinsi Jawa Tengah

Silvi Eka Wulandari\*, Siti Sunendiari

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam,  
Universitas Islam Bandung, Indonesia.

\*silviekaw@gmail.com, diarisunen22@gmail.com

**Abstract.** The Small Area Estimation (SAE) method can be used to estimate the district/city level Gross Enrollment Rate of Higher Education with a better level of precision because indirect estimation is carried out by utilizing the strength of the auxiliary variables. If there is a linearity assumption that is met and not met at the same time, a semiparametric method will be used. Penalized spline is used so that the number of knots produced are not too many. Then, to evaluate the results, it is done by calculating Mean Square Error (MSE) using the jackknife method. In this paper, it will make a model of the district/city level Gross Enrollment Rate of Higher Education in Central Java Province in 2019. From the results, it can be concluded that Gross Enrollment Rate of Higher Education modeling using the SAE method with a semiparametric penalized spline produces different models for each district/city because there are different values of  $(x_i - k_j)$ . Then the average value of MSE is 6805,79763012 where these values are smaller than the average values of MSE for direct estimation, which is 7950,17170752. Therefore, it can be said that the model formed using the SAE method with the semiparametric penalized spline has a better level of precision.

**Keywords:** Gross Enrollment Rate of Higher Education, Jackknife Method, Semiparametric Penalized Spline, Small Area Estimation.

**Abstrak.** Metode *Small Area Estimation* (SAE) digunakan untuk melakukan estimasi Angka Partisipasi Kasar Perguruan Tinggi (APK-PT) tingkat kabupaten/kota dengan tingkat presisi yang lebih baik karena dilakukannya pendugaan secara tidak langsung dengan memanfaatkan kekuatan dari variabel penyerta. Jika terdapat asumsi linearitas yang terpenuhi dan tidak terpenuhi secara sekaligus maka dilakukan pendekatan secara semiparametrik. Pendekatan *penalized spline* digunakan agar jumlah knot yang dihasilkan tidak terlampaui banyak. Kemudian, untuk mengevaluasi hasil dugaan dilakukan dengan menghitung nilai *Mean Square Error* (MSE) menggunakan metode *jackknife*. Pada penelitian ini akan dilakukan pemodelan APK-PT tingkat kabupaten/kota di Provinsi Jawa Tengah tahun 2019. Dari penelitian yang telah dilakukan, dapat disimpulkan bahwa pemodelan APK-PT menggunakan metode SAE dengan pendekatan semiparametrik *penalized spline* menghasilkan model yang berbeda-beda untuk setiap kabupaten/kota nya karena terdapat perbedaan nilai  $(x_i - k_j)$ . Kemudian diperoleh nilai rata-rata MSE yaitu 6805,79763012 dimana nilai tersebut lebih kecil jika dibandingkan dengan nilai rata-rata MSE untuk pendugaan langsung yaitu 7950,17170752. Oleh karena itu, dapat dikatakan bahwa model yang terbentuk dengan menggunakan metode SAE dengan pendekatan semiparametrik *penalized spline* memiliki tingkat presisi yang lebih baik.

**Kata Kunci:** Angka Partisipasi Kasar Perguruan Tinggi, Metode *Jackknife*, Semiparametrik *Penalized Spline*, *Small Area Estimation*.

## 1. Pendahuluan

Metode *Small Area Estimation* (SAE) merupakan metode yang digunakan untuk menduga suatu karakteristik dari subpopulasi tertentu (Sadik, 2008). Metode SAE melakukan pendugaan secara tidak langsung (*indirect estimation*) dengan berdasar kepada pemodelan yang memanfaatkan “kekuatan” dari informasi tambahan yang disebut dengan variabel penyerta (*model-based*). Berbeda dengan pendugaan secara langsung (*direct estimation*) dimana pendugaan yang dilakukan berdasarkan kepada desain penarikan sampel (*design-based*). Pendugaan secara tidak langsung ini merupakan sebuah alternatif dari pendugaan secara langsung yang dipercaya dapat meningkatkan efektivitas dari ukuran sampel yang digunakan dan dapat menurunkan tingkat *error* (Darsyah, 2013). Metode SAE dapat digunakan untuk menghasilkan dugaan dengan tingkat presisi yang lebih baik tanpa harus dilakukannya penambahan sampel (Firmando & Ubaidillah, 2019).

Metode SAE yang sering digunakan dalam penelitian ialah *Empirical Best Linear Unbiased Prediction* (EBLUP), *Empirical Bayes* (EB), dan *Hierarchical Bayes* (HB) karena memiliki dampak yang signifikan pada hasil estimasi di area kecil (Ghosh & Rao, 1994). Metode EBLUP dilakukan dengan menggantikan varians yang tidak diketahui dengan penduga varians yang diperoleh melalui data sampel sehingga dapat meminimumkan nilai *Mean Square Error* (MSE) (Apriani, 2017).

Pada penelitian ini, metode SAE akan dikembangkan dengan menggunakan pendekatan semiparametrik sehingga dapat menghasilkan model yang lebih fleksibel karena mengandung komponen yang bersifat parametrik dan nonparametrik (Ruppert *et al.*, 2003). Suatu komponen dikatakan parametrik ketika memenuhi asumsi linearitas dan dikatakan komponen nonparametrik ketika tidak memenuhi asumsi linearitas. Salah satu pendekatan semiparametrik yang digunakan adalah *penalized spline* yang merupakan suatu pendekatan pemulusan yang dikenal akan kesederhanaan dan fleksibilitasnya (Ruppert *et al.*, 2003). Evaluasi hasil pendugaan dilakukan dengan menghitung nilai MSE dengan menggunakan metode *jackknife*.

Angka Partisipasi Kasar Perguruan Tinggi (APK-PT) merupakan proporsi jumlah penduduk yang berkuliah di perguruan tinggi terhadap jumlah penduduk yang berusia kuliah (19-24 tahun) (BPS, 2020). Besarnya APK-PT menunjukkan kualitas pemerintah dalam melayani masyarakatnya agar dapat memperoleh akses untuk menempuh pendidikan tinggi (Habibah, *et al.*, 2019).

Berdasarkan latar belakang yang telah diuraikan, maka tujuan yang ingin dicapai dalam penelitian ini adalah:

1. Untuk mengetahui pemodelan APK-PT menggunakan metode SAE dengan pendekatan semiparametrik *penalized spline* di Provinsi Jawa Tengah pada tahun 2019.
2. Untuk mengetahui perbandingan antara hasil dugaan APK-PT secara langsung dan tidak langsung di Provinsi Jawa Tengah pada tahun 2019.

## 2. Metodologi

Pada penelitian ini, data yang digunakan berasal dari hasil Survey Sosial Ekonomi dan Nasional Badan Pusat Statistik tahun 2019. Unit penelitian yang digunakan adalah 33 kabupaten/kota di Provinsi Jawa Tengah pada tahun 2019 yang terdiri atas 27 kabupaten dan 6 kota. Dua kabupaten pada Provinsi Jawa Tengah yaitu Kabupaten Purbalingga dan Kabupaten Batang tidak diikutsertakan dalam penelitian ini dikarenakan adanya ketidaklengkapan pencatatan data untuk variabel penyerta yang digunakan. Variabel yang digunakan dalam penelitian ini, yaitu data APK-PT setiap kabupaten/kota yang berperan sebagai variabel respon kemudian data jumlah penduduk miskin, jumlah perguruan tinggi, rasio dosen-mahasiswa, dan rata-rata banyaknya anggota rumah tangga yang berperan sebagai variabel penyerta.

Penelitian ini akan menghasilkan pemodelan dengan menggunakan metode SAE dengan pendekatan semiparametrik *penalized spline*. Digunakan metode SAE karena metode ini melakukan pendugaan secara tidak langsung dimana hasil dugaan yang dihasilkan dipercaya memiliki tingkat presisi yang lebih baik dibandingkan dengan pendugaan secara langsung. Kemudian, digunakan pendekatan secara semiparametrik karena metode tersebut dapat

menangani komponen yang bersifat parametrik dan nonparametrik secara sekaligus. Lalu, digunakan metode *penalized spline* agar jumlah knot yang dihasilkan tidak terlampaui banyak sehingga model yang dihasilkan akan lebih fleksibel.

Berikut ini merupakan tahapan analisis yang dilakukan, yaitu:

1. Membuat statistik deskriptif untuk setiap variabel.
2. Mendeteksi adanya kasus multikolinearitas dengan melihat nilai VIF.
3. Membuat *scatterplot* antara variabel respon dan prediktor.
4. Menduga APK-PT menggunakan metode SAE dengan pendekatan semiparametrik *penalized spline* dengan tahapan sebagai berikut:
  - a. Proses pembentukan model yaitu  $y = f(X_1) + f(X_2, X_3, X_4) + u$ . Pada model tersebut  $f(\text{parametrik})$  dijelaskan oleh  $f(X_1)$  dan nilai nonparametrik *penalized spline* dijelaskan oleh  $f(X_2, X_3, X_4)$ .
  - b. Menghitung nilai lambda ( $\lambda$ ) optimum menggunakan kriteria GCV.
  - c. Menentukan jumlah knot yang digunakan dengan menggunakan *fixed selection method* serta lokasi titik knot.
  - d. Melakukan estimasi untuk  $\beta$ ,  $\gamma$ , dan  $u$ .
  - e. Melakukan pemodelan APK-PT untuk 33 kabupaten/kota di Provinsi Jawa Tengah tahun 2019.
5. Menghitung nilai MSE dari hasil dugaan APK-PT dengan menggunakan metode *jackknife*.

### 3. Pembahasan dan Diskusi

#### Analisis Deskriptif Data Penelitian

Berikut ini merupakan hasil deskripsi data untuk setiap variabel yang diolah dengan menggunakan *software* SPSS.

**Tabel 1.** Statistik Deskriptif Variabel Penelitian

Variabel	Minimum	Maksimum	Rata-rata
APK-PT	5,34	46,89	18,12
Jumlah Penduduk Miskin	9.100	293.200	107.245
Jumlah Perguruan Tinggi	1	64	8,18181818
Rasio Dosen-Mahasiswa	0,02271388	0,22842640	0,06662446
Rata-rata Banyaknya Anggota Rumah Tangga	3	4	3,81818182

Berdasarkan hasil deskripsi data diatas, dapat dijelaskan beberapa hal sebagai berikut:

1. Nilai rata-rata APK-PT untuk 33 kabupaten/kota di Provinsi Jawa Tengah adalah sebesar 18,12%. APK-PT tertinggi sebesar 46,89% yang berada di Kota Surakarta dan terendah sebesar 5,34% yang berada di Kabupaten Banjarnegara.
2. Jumlah rata-rata perguruan tinggi adalah 8 perguruan tinggi. Kota Semarang merupakan wilayah dengan jumlah perguruan tinggi terbanyak yaitu berjumlah 64 perguruan tinggi sedangkan Kabupaten Demak merupakan wilayah dengan jumlah perguruan tinggi paling sedikit yaitu hanya terdapat satu perguruan tinggi.
3. Nilai rata-rata rasio antara dosen terhadap mahasiswa sebesar 0,06662446 atau 6,662446% dimana artinya rata-rata satu orang dosen mengajar sebanyak kurang lebih

15 mahasiswa. Nilai rasio antara dosen terhadap mahasiswa tertinggi berada di Kabupaten Sragen dengan nilai sebesar 0,22842640 atau 22,842640% yang berarti satu orang dosen mengajar kurang lebih sebanyak 4 mahasiswa sedangkan Kabupaten Sukoharjo adalah wilayah dengan angka rasio antara dosen terhadap mahasiswa terendah yaitu sebesar 0,02271388 atau 2,271388% yang berarti satu orang dosen mengajar kurang lebih 44 mahasiswa. Dari hal tersebut, dapat disimpulkan bahwa semakin rendah nilai rasio antara dosen terhadap mahasiswa, maka jumlah mahasiswa yang diajar oleh satu orang dosen akan semakin banyak dan semakin tinggi nilai rasio antara dosen terhadap mahasiswa maka jumlah mahasiswa yang diajar oleh satu orang dosen akan semakin sedikit yang berarti partisipasi penduduk dalam menempuh pendidikan tinggi pun akan rendah.

4. Nilai rata-rata jumlah penduduk miskin sebanyak 107.245 jiwa dengan jumlah penduduk miskin terbanyak sebesar 292.200 jiwa yang berada di Kabupaten Brebes dan jumlah penduduk miskin paling sedikit yaitu 9.100 jiwa yang berada di Kota Magelang.
5. Nilai rata-rata banyaknya anggota rumah tangga berjumlah 4 jiwa.

### Uji Multikolinearitas

Salah satu cara untuk mengetahui apakah terjadi kasus multikolinearitas adalah dengan melihat nilai *Variance Inflation Factor* (VIF) untuk setiap variabel prediktor. Berikut ini merupakan nilai VIF yang diperoleh untuk masing-masing variabel prediktor:

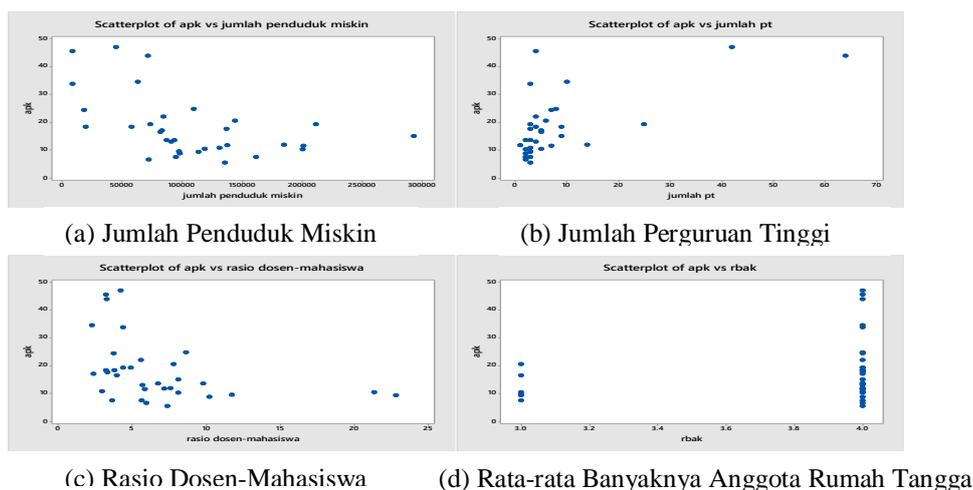
**Tabel 2.** Nilai VIF Variabel Prediktor

Variabel	VIF
X1	1,06549494
X2	1,04293869
X3	1,59780684
X4	1,49418693

Dari tabel di atas dapat diketahui bahwa nilai VIF untuk setiap variabel prediktor bernilai kurang dari 10, sehingga dapat disimpulkan bahwa tidak terdapat multikolinearitas diantara variabel prediktor pada penelitian ini.

### Pola Hubungan antara Variabel Respon dengan Variabel Prediktor

Dengan dilakukannya langkah ini, maka akan diketahui variabel mana yang bersifat parametrik dan nonparametrik. Berikut ini merupakan hasil *scatterplot* yang terbentuk antara variabel respon dengan masing-masing variabel prediktor.



**Gambar 1.** Scatterplot antara Variabel Respon dengan Variabel Prediktor

Dari gambar di atas dapat diduga bahwa variabel jumlah penduduk miskin merupakan variabel yang bersifat parametrik karena *scatterplot* yang terbentuk membentuk suatu pola kelinearan yang mengarah ke arah kiri dan variabel jumlah perguruan tinggi, rasio dosen-mahasiswa, dan rata-rata banyaknya anggota rumah tangga merupakan variabel yang bersifat nonparametrik karena *scatterplot* yang terbentuk tidak menunjukkan adanya suatu pola tertentu.

**Pemodelan APK-PT dengan Menggunakan Metode SAE dengan Pendekatan Semiparametrik *Penalized Spline***

Pada penelitian ini, terdapat tiga variabel yang memiliki pola hubungan non-linear yaitu  $x_2, x_3$ , dan  $x_4$  serta terdapat satu variabel yang memiliki pola hubungan linear yaitu variabel  $x_1$  sehingga model dapat ditulis sebagai berikut:

$$\hat{y} = f(x_1) + f(x_2, x_3, x_4) + u \quad \dots \quad (1)$$

Untuk mendapatkan nilai  $f(x_2, x_3, x_4)$  dibutuhkan perhitungan untuk mendapatkan nilai knot terlebih dahulu karena ketiga variabel tersebut bersifat nonparametrik tidak seperti  $x_1$  yang bersifat parametrik sehingga untuk mendapatkan nilai  $f(x_1)$  tidak perlu menggunakan nilai knot.

Penentuan banyaknya jumlah knot yang digunakan dilakukan dengan rumus *fixed selection method* yang memberikan hasil bahwa jumlah knot yang digunakan untuk variabel  $x_2$  sebanyak 3 knot, untuk variabel  $x_3$  sebanyak 8 knot, dan untuk variabel  $x_4$  sebanyak 1 knot. Berikut ini merupakan hasil perhitungan nilai knot yang digunakan untuk variabel  $x_2, x_3$ , dan  $x_4$  dengan menggunakan bantuan *software* R-Studio.

**Tabel 3.** Nilai Knot untuk Variabel  $x_2, x_3$ , dan  $x_4$

Variabel	Nilai Knot
$x_2$	3,88888900, 6,77777800, 9,66666700.
$x_3$	0,03226655, 0,03592483, 0,04047595, 0,04827616, 0,05745934, 0,07079375, 0,07852235, 0,09410926.
$x_4$	3,66666700

Setelah diketahui nilai knot untuk variabel  $x_2, x_3$ , dan  $x_4$ , maka model yang terbentuk menjadi seperti berikut:

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \gamma_1(x_2 - 3,88888900) + \gamma_2(x_2 - 6,77777800) + \gamma_3(x_2 - 9,66666700) + \beta_3 x_3 + \gamma_4(x_3 - 0,03226655) + \gamma_5(x_3 - 0,03592483) + \gamma_6(x_3 - 0,04047595) + \gamma_7(x_3 - 0,04827616) + \gamma_8(x_3 - 0,05745934) + \gamma_9(x_3 - 0,07079375) + \gamma_{10}(x_3 - 0,07852235) + \gamma_{11}(x_3 - 0,09410926) + \beta_4 x_4 + \gamma_{12}(x_4 - 3,66666700) + u \quad \dots (2)$$

Persamaan (2) merupakan persamaan dengan memasukkan nilai knot yang ditampilkan pada Tabel. 3. Oleh karena itu, untuk menyempurnakan pemodelan yang terbentuk maka dibutuhkan nilai  $\hat{\beta}$  yang diperoleh dengan memaksimalkan fungsi *likelihood*. Berikut ini perolehan nilai estimasi  $\beta$  yang disajikan pada tabel berikut:

**Tabel 4.** Nilai Estimasi  $\beta$

Parameter	Nilai Estimasi
$\hat{\beta}_0$	21,15845828
$\hat{\beta}_1$	-0,00007530
$\hat{\beta}_2$	0,53113305
$\hat{\beta}_3$	-38,03342223
$\hat{\beta}_4$	0,86360754

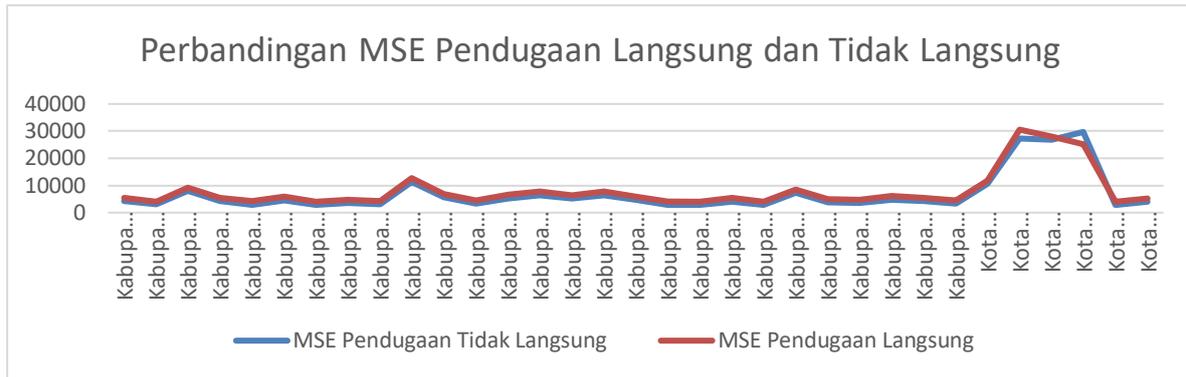
Setelah diketahui nilai estimasi  $\beta$  dan nilai knot, maka model yang akan digunakan untuk menduga APK-PT di 33 kabupaten/kota Provinsi Jawa Tengah tahun 2019 adalah sebagai berikut:

$$\hat{y} = 21,15845828 - 0,00007530x_1 + 0,5311330x_2 + \gamma_1(x_2 - 3,88888900) + \gamma_2(x_2 - 6,77777800) + \gamma_3(x_2 - 9,66666700) - 38,03342223x_3 + \gamma_4(x_3 - 0,03226655) + \gamma_5(x_3 - 0,03592483) + \gamma_6(x_3 - 0,04047595) + \gamma_7(x_3 - 0,04827616) + \gamma_8(x_3 - 0,05745934) + \gamma_9(x_3 - 0,07079375) + \gamma_{10}(x_3 - 0,07852235) + \gamma_{11}(x_3 - 0,09410926) + 0,86360754x_4 + \gamma_{12}(x_4 - 3,66666700) + u \quad \dots (3)$$

Model yang terbentuk memiliki efek acak yang dinotasikan oleh  $\gamma$  dan  $u$ . Pada penelitian ini komponen varians untuk efek acak ditaksir melalui metode *restricted maximum likelihood* karena jika penaksiran dilakukan dengan menggunakan *maximum likelihood* akan menghasilkan dugaan yang bersifat bias. Prediktor  $\beta, \gamma$  dan  $u$  merupakan *empirical best linear unbiased predictor* (EBLUP) karena mensubstitusikan komponen varians yang tidak diketahui ke dalam penduga BLUP.

#### **Perhitungan MSE Menggunakan Metode Jackknife**

Setelah mendapatkan nilai dugaan APK-PT untuk 33 kabupaten/kota di Provinsi Jawa Tengah pada tahun 2019, langkah selanjutnya adalah menghitung nilai MSE baik untuk pendugaan langsung dan pendugaan tidak langsung. Berikut ini merupakan visualisasi untuk perbandingan MSE pendugaan langsung dan tidak langsung.



**Gambar 2.** Perbandingan MSE Pendugaan Langsung dan Tidak Langsung

Dari gambar di atas, terlihat bahwa nilai MSE pendugaan langsung dan tidak langsung memiliki nilai yang tidak berbeda jauh namun nilai MSE pendugaan tidak langsung relatif berada di bawah garis nilai MSE pendugaan langsung. Oleh karena itu, dapat disimpulkan bahwa pendugaan APK-PT di 33 kabupaten/kota Provinsi Jawa Tengah pada tahun 2019 menggunakan metode SAE dengan pendekatan semiparametrik *penalized spline* memiliki tingkat presisi yang lebih baik dibandingkan pendugaan yang dilakukan secara langsung karena memiliki nilai MSE yang lebih kecil.

**4. Kesimpulan**

Berdasarkan hasil analisis dan pembahasan yang telah dilakukan, peneliti menyimpulkan beberapa hasil penelitian sebagai berikut:

1. Pemodelan APK-PT di 33 kabupaten/kota Provinsi Jawa Tengah pada tahun 2019 menggunakan metode SAE berbasis area dengan pendekatan semiparametrik *penalized spline* menghasilkan pemodelan yang berbeda-beda karena terdapat perbedaan nilai  $(x_i - k_j)$  untuk setiap kabupaten/kota nya.
2. Hasil dugaan APK-PT secara tidak langsung di 33 kabupaten/kota Provinsi Jawa Tengah pada tahun 2019 memiliki tingkat presisi yang lebih baik karena diperolehnya nilai rata-rata MSE yaitu 6805,79763012 dimana nilai tersebut lebih kecil jika dibandingkan dengan nilai rata-rata MSE untuk hasil dugaan langsung yaitu 7950,17170752.

**Acknowledge**

Penulis mengucapkan banyak terima kasih kepada seluruh pihak yang telah membantu untuk terlaksananya penelitian ini, khususnya kepada Ibu Siti Sunendiari, Dra., M.Si. yang selalu memberikan bantuan dan saran kepada penulis serta para dosen Statistika Unisba yang telah memberikan banyak ilmu pengetahuan kepada penulis. Tak lupa juga penulis ucapkan terima kasih kepada rekan-rekan yang selalu membantu penulis untuk saling bertukar pikiran hingga akhirnya penelitian ini dapat terselesaikan dengan baik.

**Daftar Pustaka**

[1] Apriani, F. (2017). *Pemodelan Pengeluaran Per Kapita Menggunakan Small Area Estimation dengan Pendekatan Semiparametrik Penalized Spline*. Tesis. Program Magister, Program Studi Matematika, Institut Teknologi Sepuluh Nopember.

[2] Badan Pusat Statistik. (2020). *Statistik Pendidikan Provinsi Jawa Tengah Tahun 2019*. Semarang: BPS.

[3] Darsyah, M. Y. (2013). Small Area Estimation terhadap Pengeluaran per Kapita di Kabupaten Sumenep dengan Pendekatan Nonparametrik. *Statistika*, **1**(2), 28-36.

[4] Firmando, G. & Ubaidillah, A. (2019). *Pendugaan Area Kecil untuk Angka Partisipasi Kasar Pendidikan Dasar dan Menengah Tingkat Kabupaten/Kota di Provinsi Jawa Tengah*

*tahun 2018*. Makalah dipresentasikan dalam Seminar Nasional Official Statistics 2019: Pengembangan Official Statistics dalam mendukung Implementasi SDG's, Politeknik Statistika STIS, Jakarta, 24 September 2019.

- [5] Ghosh, M. & Rao, J. N. K. (1994). Small Area Estimation: an Appraisal. *Statistical Science*, **9**(1), 55-93.
- [6] Habibah, S., Putra, Y. P., & Putra, Y. M. (2019). Faktor-faktor yang Mempengaruhi Angka Partisipasi Perguruan Tinggi pada 32 Provinsi di Indonesia tahun 2013-2016. *Jurnal Anggaran dan Keuangan Negara Indonesia*, **1**(1), 17-21.
- [7] Ruppert, D., Wand, M. P., & Carrol. (2003). *Semiparametric Regression*. United Kingdom: Cambridge University.
- [8] Sadik, K. (2008). Parameter *Quantile-like* dalam Pendugaan Area Kecil Melalui Pendekatan *Penalized Splines*. *Statistika*, **8**(1), 31-36.
- [9] Shofwani Sheila Ghazia, Kudus Abdul. (2021). *Penentuan Kriteria Pengunjung dalam Pemilihan Green Hotel di Kota Bandung Menggunakan Metode Discrete Choice Experiment dengan Desain Choice Sets Kombinatorial*. *Jurnal Riset Statistika*, **1**(1), 1-9.