**TESIS – SS 142501**

# PEMODELAN PENGELUARAN PER KAPITA MENGGUNAKAN SMALL AREA ESTIMATION DENGAN PENDEKATAN SEMIPARAMETRIK PENALIZED SPLINE

FARIDA APRIANI
NRP. 1315201026

DOSEN PEMBIMBING :

Dr.Agnes Tuti Rumiati, M.Sc.
Dr. Wahyu Wibowo, M.Si

PROGRAM MAGISTER
JURUSAN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA
2017

THESIS – SS 142501

# SEMIPARAMETRIC SMALL AREA ESTIMATION USING PENALIZED SPLINE TO MODELING PER CAPITA EXPENDITURE IN SLEMAN

FARIDA APRIANI
NRP. 1315201026

SUPERVISOR :

Dr.Agnes Tuti Rumiati, M.Sc.
Dr. Wahyu Wibowo, M.Si

PROGRAM OF MAGISTER
DEPARTMENT OF STATISTICS
FACULTY OF MATHEMATICS AND NATURAL SCIENCE
SEPULUH NOPEMBER INSTITUTE OF TECHNOLOGY
SURABAYA
2017

# PEMODELAN PENGELUARAN PER KAPITA MENGGUNAKAN SMALL AREA ESTIMATION DENGAN PENDEKATAN SEMIPARAMETRIK PENALIZED SPLINE

Tesis disusun untuk memenuhi salah satu syarat memperoleh gelar Magister Sains (M.Si)
di
Institut Teknologi Sepuluh Nopember
Oleh:

**FARIDA APRIANI**
**NRP. 1315 201 026**

Tanggal Ujian : 16 Januari 2017
Periode Wisuda : Maret 2017

Disetujui oleh:

1. Dr. Agnes Tuti Rumiati, M.Sc (Pembimbing I)
NIP. 19570724 198503 2 002

2. Dr. Wahyu Wibowo,M.Si (Pembimbing II)
NIP. 19740328 199802 1 001

3. Dr. rer.pol. Dedy Dwi P, M.Si (Penguji)
NIP. 19831204 200812 1 002

4. Dr. Vita Ratnasari, M.Si (Penguji)
NIP. 19700910 199702 2 001

an. Direktur Program Pascasarjana
Asisten Direktur

Prof. Dr. Ir. Tri Widjaja, M.Eng.
NIP. 19611021 198603 1 001

Direktur Program Pasca Sarjana,

Prof. Ir. Djauhar Manfaat, M.Sc., Ph.D.
NIP.19601202 198701 1 001

# PEMODELAN PENGELUARAN PER KAPITA MENGGUNAKAN SMALL AREA ESTIMATION DENGAN PENDEKATAN SEMIPARAMETRIK PENALIZED SPLINE

**Nama Mahasiswa : Farida Apriani**
**NRP : 1315201026**
**Pembimbing : Dr. Agnes Tuti Rumiati, M.Sc.**
**Co- Pembimbing : Dr. Wahyu Wibowo, M.Si**

## ABSTRAK

Kebutuhan informasi untuk domain dan wilayah yang lebih kecil semakin diminati. Keterbatasan sampel pada kegiatan survei menjadi kendala untuk menyediakan data untuk domain dan wilayah yang lebih kecil. *Small Area Estimation* (SAE) dapat dilakukan untuk mengatasi permasalahan ini. Teknik pendugaan ini memanfaatkan data dari domain besar untuk menduga parameter pada domain yang lebih kecil yang dapat berupa desa/kelurahan, kecamatan, kabupaten, kelompok suku, maupun kelompok umur. Pada penelitian ini akan menerapkan pemodelan semiparametrik *penalized spline* untuk menduga area kecil level kecamatan di Kabupaten Sleman., data yang digunakan pada penelitian ini adalah data yang berasal dari SUSENAS dan PODES Kabupaten Sleman tahun 2014, didapatkan hasil rata-rata pengeluaran per kapita tahun 2014 dengan menggunakan metode estimasi tidak langsung adalah sebesar Rp 846.891,5 dengan pengeluaran per kapita tertinggi terdapat pada kecamatan Ngaglik.

**Kata Kunci**— Pengeluaran per Kapita, Semiparametrik *Penalized Spline*, *Small Area Estimation*, Sleman.

# Semiparametric Small Area Estimation Using Penalized Spline to Modeling Per Capita Expenditure In Sleman

**Nama Mahasiswa : Farida Apriani**
**NRP : 1315201026**
**Supervisor : Dr. Agnes Tuti Rumiati, M.Sc.**

**Co- Supervisor : Dr. Wahyu Wibowo, M.Si**

**ABSTRACT**

Information for domain and smaller area is more attractive but limitation of sample in this survey become obstacle to provide data for smaller area. Small area estimation can overcame this problem, this method is statistical technique used to estimate parameters of a subpopulation that small sample size and this method utilizing data from a large domain to estimate smaller domain that can be either rural/village, district, county, or tribal group. This study will apply modelling of small area estimation using semiparametric penalized spline approach at district level in sleman. In this study, data used came from SUSENAS and PODES sleman district in 2014. Result from this study, average of expenditure per capita in sleman in 2014 was Rp 846.891,5 and the highest expenditure per capita in Ngaglik district

**Keywords**— Expenditure Per Capita , Semiparametrik *Penalized Spline, Small Area Estimation,* Sleman.

# KATA PENGANTAR

Puji Syukur yang sebesarnya penulis panjatkan kehadirat Allah SWT atas segala limpahan rahmat dan karunia-Nya kepada penulis sehingga dapat menyelesaikan penulisan tesis dengan judul ***PEMODELAN PENGELUARAN PER KAPITA MENGGUNAKAN SMALL AREA ESTIMATION DENGAN PENDEKATAN SEMIPARAMETRIK PENALIZED SPLINE*** dengan baik.

Tesis ini disusun untuk memenuhi salah satu syarat memperoleh gelar Magister Sains (M.Si) di Jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Teknologi Sepuluh Nopember, Surabaya.

Begitu banyak bantuan dari berbagai pihak hingga tesis ini dapat diselesaikan dengan baik dan lancer. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan terima kasih yang sedalamnya kepada yang terhormat:

1. Bapak Dr. Suhartono, M.Sc., selaku Ketua Jurusan Statistika yang telah memberikan banyak dukungan dalam penyelesaian tesis mahasiswa dan mahasiswi S2 Statistika
2. Bapak Dr. rer.pol. Heri Kuswanto, M.Si selaku ketua Program studi S2 Statistika yang telah memberi bimbingan dan dukungan kepada mahasiswa selama kuliah hingga mampu menyelesaikan tesis.
3. Ibu. Irhamah, M.Si, Ph.D. selaku dosen pembimbing akademik yang telah memberikan masukkan dan bimbingan selama perkuliahan
4. Ibu Dr. Agnes Tuti Rumiati, M.Sc. dan Bapak Dr. Wahyu Wibowo, M.Si. selaku dosen pembimbing dalam penyusunan tesis ini, yang telah banyak memberikan saran dan masukkan dengan penuh kesabaran serta bersedia meluangkan waktu untuk membimbing penulis.
5. Bapak Dr.rer.pol. Dedy Dwi P, M.Si.dan Ibu Dr.Vita Ratnasari, M.Si. selaku dosen penguji yang telah banyak memberi saran dan masukan untuk perbaikan tesis ini.

6. Kedua Orangtua tercinta, Ayah Kusnan dan Ibu Ema , terima kasih atas segala doa dan dukungan tiada henti, semoga pencapaian ini memberikan setitik kebahagiaan untuk emak dan ayah. Terima kasih juga kepada kakak-kakak tercinta yang telah mendukung dan menyemangati adik bungsunya ini, serta sanak keluarga lainya yang senantiasa mendoakan dan memberikan motivasi untuk kesuksesan penulis.
7. Bapak dan Ibu dosen Jurusan Statistika FMIPA ITS yang telah memberikan ilmu yang bermanfaat, serta seluruh staf administrasi akademik, laboratorium, dan ruang baca statistika FMIPA ITS yang telah memberikan pelayanan fasilitas selama perkuliahan.
8. Pascasarjana ITS yang telah memberi kesempatan kepada penulis untuk dapat kuliah di salah satu universitas terbaik di Indonesia
9. Bapak dan Ibu Dosen Program Studi Statistika FMIPA UII, almamater penulis yang telah memberikan ilmu yang bermanfaat serta selalu memberikan semangat dalam menyelesaikan penulisan tesis ini.
10. Teman-teman seperjuangan Magister Statistika angkatan 2015 serta teman-teman “cewek-cewek sholehah” terima kasih atas saran, kerjasama dan kebersamaannya.

Penulis menyadari bahwa masih terdapat banyak kekurangan dalam penulisan tesis ini. Oleh karena itu, penulis berharap pembaca dapat memberikan kritik dan saran yang membangun untuk perbaikan materi atau isi dari tesis ini, dan semoga hasil penelitian ini dapat bermanfaat bagi perkembangan ilmu pengetahuan.

Surabaya, Desember 2016

Penulis

# DAFTAR ISI

**Halaman**

**HALAMAN JUDUL** ... i
**HALAMAN PENGESAHAN** ... iii
**ABSTRAK** ... iv
**ABSTRACT** ... v
**KATA PENGANTAR** ... vi
**DAFTAR ISI** ... viii
**DAFTAR TABEL** ... x
**DAFTAR GAMBAR** ... xi
**DAFTAR LAMPIRAN** ... xii
**BAB I PENDAHULUAN** ... 1
1.1 Latar Belakang ... 1
1.2 Perumusan masalah ... 4
1.3 Tujuan Penelitian ... 5
1.4 Manfaat Penelitian ... 5
1.5 Batasan Masalah ... 5
**BAB II TINJAUAN PUSTAKA** ... 7
2.1 *Small Area Estimation* ... 7
2.1.1 Model Berbasis Area ... 8
2.1.2 Model Berbasis Unit ... 9
2.2 Model Regresi *Spline* ... 10
2.3 Model Regresi *Penalized Spline* ... 11
2.4 Pendugaan Area Kecil dengan Semiparametrik *Penalized Spline* ... 14
2.5 *Best Linear Unbiased Prediction* (BLUP) *dan Empirical Best Linear Unbiased Prediction* (EBLUP) ... 16
2.6 Pendugaan MSE dengan Menggunakan Metode *Jackknife* ... 18
2.7 Survei Sosial Ekonomi Nasional (SUSENAS) ... 20
2.8 Pengeluaran Perkapita ... 23
**BAB III METODOLOGI PENELITIAN** ... 25
3.1 Sumber Data ... 25
3.2 Variabel Penelitian ... 25
3.3 Langkah Analisis Model Pendugaan Area Kecil Menggunakan Regresi Semiparametrik *Penalized Spline* ... 26

3.4 Diagram Alir Penelitian ........ 28
**BAB IV HASIL DAN PEMBAHASAN ........ 29**
4.1 Estimator *Small Area Estimation* dengan Pendekatan Semiparametrik *Penalized* Spline ........ 29
4.2 Aplikasi Pemodelan Pengeluaran Per Kapita dengan *Small Area Estimation* dengan Semiparametrik *Penalized* Spline ........ 36
4.2.1 Deskripsi Umum Kabupaten Sleman ........ 36
4.2.2 Eksplorasi Data Kabupaten Sleman ........ 36
4.2.3 Eksplorasi Data Hubungan Antar Variabel ........ 39
4.2.4 Pemodelan Pengeluaran Per Kapita dengan Menggunakan Semiparametrik *Penalized* Spline ........ 42
4.3 Kebaikan Model Pendugaan ........ 48
4.4 Pembahasan *Small Area Estimation* dengan Semiparametrik *Penalized* Spline ........ 51
**BAB V KESIMPULAN DAN SARAN ........ 53**
5.1 Kesimpulan ........ 53
5.2 Saran ........ 54
**DAFTAR PUSTAKA** ........ 55
**LAMPIRAN** ........ 59

# DAFTAR TABEL

**Halaman**

Tabel 2.1a Daftar Pengambilan Sampel pada Susenas Kabupaten Sleman .. 22

Tabel 2.1b Daftar Pengambilan Sampel pada Susenas Kabupaten Sleman .. 23

Tabel 3.1 Variabel Penelitian ........................................................................ 25

Tabel 4.1 Nilai estimasi $\beta$ ........................................................................ 44

Tabel 4.2 Model yang terbentuk untuk setiap *Small Area Estimation* .......... 46

Tabel 4.3 Ringkasan Pendugaan Pengeluaran Per Kapita di Kabupaten Sleman ........................................................................................... 47

Tabel 4.4 Hasil Pendugaan Pengeluaran Per Kapita Kabupaten Sleman Level Kecamatan 2014 ........................................................................ 48

Tabel 4.5 Hasil MSE dengan *Jackknife* .......................................................... 49

Tabel 4.5 Ringkasan MSE *Jackknife* .............................................................. 50

# DAFTAR GAMBAR

**Halaman**


Gambar 2.1 Ilustrasi Susenas ........................................................................ 22

Gambar 3.3 Diagram Alir Penelitian ............................................................ 28

Gambar 4.1 Kabupaten Sleman...................................................................... 36

Gambar 4.2 Kepadatan Penduduk Kabupaten Sleman 2014 .......................... 37

Gambar 4.3 Persentase Petani Kabupaten Sleman 2014................................ 38

Gambar 4.4 Penduduk Sekolah di Kabupaten Sleman 2014 .......................... 38

Gambar 4.5 Jumlah Pengguna Listrik di Kabupaten Sleman 2014 ................ 39

Gambar 4.6 Scatterplot Empat Variabel ........................................................ 40

Gambar 4.7 Scatterplot Kepadatan Penduduk ................................................ 41

Gambar 4.8 Scatterplot antara residual $f_1$ dan $x_5$ .......................................... 42

Gambar 4.9 Lokasi Titik Knot ........................................................................ 45

Gambar 4.10 Grafik MSE Pendugaan Langsung dan Tidak Langsung .......... 49

Gambar 4.11 RMSE Pendugaan langsung dan Tidak Langsung .................... 51

# DAFTAR LAMPIRAN

**Halaman**

Lampiran 1 Data Penelitian ................................................. ....................59

Lampiran 2 Data Variabel Penyerta Tingkat Kecamatan di Kabupaten Sleman BerdasarkanSleman Dalam Angka 2015, Kecamatan Dalam Angka 2015 dan Potensi Desa 2014 .............................. 60

Lampiran 3 Hasil Dugaan dan Nilai Error untuk Variabel Parametrik ...................................................................... ....................61

Lampiran 4 Nilai Estimasi Efek Acak *Penalized* Spline dengan 4 Titik Knot.. ........................................................................................ 62

Lampiran 5 RMSE Pengeluaran Per Kapita dengan Menggunakan Metode *Jackknife* ..................................................................... 63

Lampiran 6 Program Mencari Nilai Parametrik......... .................................. 64

Lampiran 7 Program Pencarian Model Regresi Nonparametrik Menggunakan *Penalized* Spline......................................................................... 65

Lampiran 8 Program Pendugaan *Small Area Estimation* dengan Pendekatan Semiparametrik........... ............................................................. 66

Lampiran 9 Program MSE *Jackknife*............................................................ 67

# BAB I

# PENDAHULUAN

## 1.1 Latar Belakang

Pembangunan merupakan suatu langkah yang dimaksudkan untuk melakukan perubahan secara struktural melalui upaya yang sistematis. Sasaran dasar pembangunan adalah penguasaan atas sumber daya manusia dan peningkatan pendidikan serta adanya partisipasi masyarakat dalam meningkatkan perekonomian suatu daerah. Dari sasaran dasar pembangunan secara ekonomi, indikator yang banyak digunakan untuk mengukur tingkat kemakmuran dan kesejahteraan adalah pendapatan per kapita. Analisis tingkat pendapatan per kapita rumah tangga sangat diperlukan oleh pemerintah dalam rangka perumusan, pelaksanaan dan evaluasi kebijakan untuk pencapaian tujuan pembangunan.

Akan tetapi pengukuran pendapatan rumah tangga secara akurat umumnya sangat sulit terutama untuk negara-negara yang sedang berkembang seperti Indonesia. Pada dasarnya pendapatan dan pengeluaran rumah tangga bukan suatu hal yang sama. Namun hubungan diantaranya sangat kuat sehingga pendekatan pola pengeluaran rumah tangga secara luas banyak digunakan untuk menganalisis pola pendapatan rumah tangga. Selain itu, ukuran pengeluaran lebih dapat dipercaya sebagai indikator pendapatan permanen rumah tangga dibandingkan dengan pendapatan. Hal ini disebabkan karena pengeluaran tidak banyak berfluktuasi dalam waktu yang singkat dibandingkan dengan pendapatan (Akita dan Pirmansyah, 2011).

Pengeluaran per kapita merupakan rata-rata besarnya pengeluaran setiap anggota keluarga. Pengeluaran rumah tangga per kapita yang terdiri dari pengeluaran makanan dan bukan makanan dapat menggambarkan bagaimana penduduk mengalokasikan kebutuhan rumah tangganya. Informasi mengenai pengeluaran per kapita ini berasal dari Survei Sosial Ekonomi Nasional (SUSENAS) yang diselenggarakan oleh BPS, dimana survei ini dirancang untuk mengumpulkan data sosial kependudukan pada lingkup yang relatif luas yaitu pada tingkat kabupaten, namun jika data hasil SUSENAS digunakan secara langsung untuk mengestimasi unit yang lebih kecil maka akan menghasilkan standar error yang cukup besar, sehingga hasil pendugaan indikator kurang dipercaya.

Upaya yang dapat dilakukan untuk menduga area kecil adalah dengan menambah sampel, namun hal ini membutuhkan biaya yang banyak sehingga untuk mengatasi masalah tersebut yaitu dengan mengoptimalkan data yang tersedia dengan menggunakan *Small Area Estimation* (SAE)

SAE merupakan suatu teknik statistika yang digunakan 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 yang dapat berupa desa/kelurahan, kecamatan, kabupaten, kelompok suku, maupun kelompok umur.  Metode SAE mempunyai konsep dalam pendugaan parameter secara tidak langsung di suatu area yang relatif kecil dalam percontohan survey (*survey sampling*) dimana pendugaan langsung tidak mampu memberikan ketelitian yang cukup bila ukuran sampel dalam *small area* berukuran kecil/sedikit, sehingga statistik yang dihasilkan akan memiliki varians yang besar atau bahkan pendugaan tidak dapat dilakukan karena tidak terwakili dalam survei (Prasad dan Rao, 1990). Pendugaan dalam SAE didasarkan pada model area kecil yang membutuhkan informasi tambahan yang memiliki hubungan dengan peubah yang sedang diamati yang disebut juga sebagai variabel penyerta *(auxiliary variabel*). Variabel penyerta ini dapat diperoleh dari survei yang lain dan diharapkan memiliki korelasi dengan peubah yang diamati (Rumiati, 2012).

Menurut Rao (2003), penggunaan estimasi tidak langsung melalui SAE dapat memberikan beberapa keuntungan yaitu (1) diagnostik model dapat digunakan untuk mendeteksi kecocokan dengan data, misalkan menggunakan analisis sisaan; (2) pengukuran presisi spesifik area dapat diasosiasikan dengan setiap estimasi pada setiap area kecil; (3) model linier campuran dengan pengaruh acak area spesifik tetap dapat dilakukan, demikian juga struktur data yang cukp kompleks misalkan struktur data *time series* atau spasial; (4) pengembangan metode untuk model pengaruh acak dapat dimanfaatkan untuk mencapai akurasi dalam area kecil.

Model SAE dapat berupa model yang berbasis area maupun berbasis unit. Penggunaan model SAE berbasis level unit telah dilakukan antara lain oleh Elber dan Lanjouw (2004), Kurnia dan Notodiputro (2006), Scealy (2010), Rumiati (2012), dan Wulansari (2013). Metode SAE berbasis area juga telah banyak dilakukan diantara lain oleh Opsomer (2004), Fausi (2011), Giusti, et al (2012) Darsyah (2013), Baskara (2014), dan Satriya (2016).

Pengembangan metode SAE dalam penelitian biasanya menggunakan pendekatan bayes atau parametrik, akan tetapi pada penelitian ini SAE akan menggunakan pendekatan semiparametrik. Metode SAE dengan menggunakan pendekatan semiparametrik mempunyai model yang lebih fleksibel dengan model linear karena keberadaan dua komponen akan mengakomodasi hubungan antara dua respon dengan prediktor yang bersifat linear, dan hubungan antar respon dengan prediktor yang bersifat nonlinear. Metode SAE yang menggunakan pendekatan semiparametrik digunakan oleh Pratesi et al (2008) dan Giusti et al (2012).

Metode semiparametrik yang akan digunakan dalam SAE ini menggunakan *penalized spline*. Menurut Hall dan Opsomer (2005), *penalized spline* merupakan satu pendekatan *smoothing* yang popular karena kesederhanaannya dan fleksibilitasnya. *Penalized spline* juga mempunyai kelebihan dibandingkan dengan metode spline lainnya yaitu dapat mengatasi model yang *overfitting* jika jumlah knot yang digunakan terlampau banyak dengan menambahkan penalty/kendala pada parameter spline dengan tujuan untuk menghindari kelebihan knot. Penggunaan metode semiparametrik dengan pendekatan *penalized spline* menurut Giusti et al (2012) dapat menangani bentuk fungsional dari hubungan antara variabel respon dan prediktor yang tidak diketahui.

Metode SAE yang sering digunakan antara lain *empirical best linear unbiased prediction* (EBLUP), *empirical bayes* (EB) dan *hierarchical bayes* (HB) *estimation* (Gosh dan Rao, 1994). Metode EBLUP dengan pendekatan *linear mixed model* (LMM) sering digunakan dalam SAE untuk variabel respon yang bersifat kontinu dan telah diketahui mempunyai efisiensi yang baik dalam SAE (Chandra, Chambers, dan Salvani, 2009). Metode EBLUP merupakan pendugaan parameter yang meminimumkan *mean square error* (MSE) dengan mensubstitusikan komponen varians yang tidak diketahui dengan penduga varian melalui data sampel

Pada penelitian ini metode SAE yang digunakan karena metode ini dapat mengestimasi sampai tingkat agregasi yang lebih rendah, sehingga metode SAE dianggap paling tepat pada penelitian ini. Dalam penelitian ini menggunakan model berbasis level area, sehingga SAE sangat membantu dalam pengestimasian model. Metode SAE digunakan untuk menduga pengeluaran per kapita di Kabupaten Sleman dimana selama periode 2010 hingga 2012, Badan Pusat Statistik (BPS) mencatat Kabupaten Sleman mengalami kenaikan pengeluaran per kapita sebesar 10,17% per tahun yaitu dari Rp. 20.813.980 pada tahun 2010 menjadi Rp. 25.260.699 pada tahun 2012. Peningkatan pengeluaran per kapita tersebut antara lain disebabkan oleh peningkatan jumlah penduduk di Kabupaten Sleman yang mengakibatkan meningkatnya konsumsi barang dan jasa. Oleh karena itu penulis tertarik untuk meneliti guna untuk menduga pengeluaran per kapita pada level kecamatan di Kabupaten Sleman dengan metode semiparametrik *penalized spline*

Metode evaluasi hasil pendugaan dilakukan dengan menduga nilai MSE *jackknife* pendugaan langsung dengan MSE jackknife pendugaan tidak langsung. Metode *jackknife* digunakan untuk memperoleh estimasi dengan tingkat presisi yang tinggi pada area kecil dengan menggunakan proses *resampling* dimana hasil yang akan didapatkan akan mendekati populasi.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang diatas, rumusan masalah dalam penelitian ini adalah

1. Bagaimana memperoleh estimasi parameter dari model SAE berbasis area menggunakan pendekatan semiparametrik *penalized spline*
2. Bagaimana menduga pengeluaran per kapita penduduk pada level kecamatan di Kabupaten Sleman menggunakan pendekatan semiparametrik *penalized spline*

## 1.3 Tujuan Penelitian

Tujuan penelitian ini adalah

1. Memperoleh estimasi parameter dari model SAE berbasis area menggunakan pendekatan semiparametrik *penalized spline*
2. Mengaplikasikan model SAE berbasis area untuk menduga pengeluaran per kapita di Kabupaten Sleman dengan pendekatan semiparametrik *penalized spline*

## 1.4 Manfaat Penelitian

Manfaat yang diperoleh dari penelitian ini adalah

1. Memberikan alternatif metode untuk mengestimasi pengeluaran per kapita pada tingkat kecamatan di Kabupaten Sleman melalui pendekatan SAE
2. Dapat dijadikan sebagai masukkan bagi pemerintah Kabupaten Sleman untuk meningkatkan perekonomian di Kabupaten Sleman berdasarkan pengeluaran per kapita

## 1.5 Batasan Masalah

Model pendugaan SAE yang digunakan merupakan model berbasis area dengan studi kasus di wilayah Kabupaten Sleman. Penelitian ini menggunakan model semiparametrik *penalized spline* dengan menggunakan variabel penyerta kepadatan penduduk, jumlah keluarga pertanian, jumlah keluarga pengguna listrik PLN, jumlah penduduk yang sedang sekolah, dan rata-rata anggota keluarga.

*(halaman ini sengaja dikosongkan)*

.

# BAB II

# TINJAUAN PUSTAKA

## 2.1. *Small Area Estimation*

Berbagai survey umumnya dirancang untuk menduga parameter populasi untuk area yang besar, seperti level nasional atau propinsi dimana pendugaan parameternya didasarkan pada desain sampling. Hal ini menyebabkan umumnya jumlah sampel kurang/tidak mencukupi untuk menghasilkan penduga langsung (*direct estimation*) yang akurat untuk pendugaan area kecil, meskipun pendugaan secara langsung ini mempunyai sifat yang tidak bias. Untuk menghadapi masalah ini diperlukan penggunaan data tambahan (seperti data sensus) untuk mendapatkan penduga yang akurat atau dapat dipercaya melalui suatu model tertentu. Pendugaan seperti ini disebut juga pendugaan tidak langsung (*indirect estimation*), dalam arti bahwa dugaan tersebut mencakup data dari domain yang lain atau dapat dikatakan bahwa pendugaan dengan menghubungkan informasi pada daerah tersebut dengan informasi dari luar area ataupun dari luar survei dengan model yang sesuai

Terdapat dua ide utama yang digunakan dalam mengembangkan metode SAE seperti yang dikembangkan oleh Fay & Heriot (1979) yaitu:

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

Menurut Rao (2003), penggunaan model SAE memberikan beberapa keuntungan yaitu:

1. Diagnostik model dapat digunakan untuk mendeteksi kecocokan dengan data, misalkan menggunakan analisis sisaan.
2. Pengukuran presisi spesifik area dapat diasosiasikan dengan setiap pendugaan area kecil

3. Model linear campuran dengan pengaruh acak area-spesifik tetap dapat dilakukan, demikian juga untuk struktur data yang cukup kompleks misalkan data time series atau spasial
4. Pengembangan metode untuk model pengaruh acak dapat dimanfaatkan untuk mencapai akurasi dalam area kecil.

SAE memiliki dua jenis model dasar yaitu model level area dasar (*basic area level model*) dan model level unit dasar (*basic unit level model*) (Rao, 2003).

### 2.1.1 Model Berbasis Area

Model berbasis area, yaitu model yang didasarkan pada ketersediaan data pendukung terdapat untuk level area tertentu. Model linear yang menjelaskan hubungan tersebut adalah sebagai berikut:

$$\theta_i = \boldsymbol{x}_i^{\boldsymbol{T}} \boldsymbol{\alpha} + b_i v_i, \quad i = 1, \cdots, m \tag{2.1}$$

Dimana $\boldsymbol{x}_i = (\boldsymbol{x}_{1i,} \boldsymbol{x}_{2i,\dots} \boldsymbol{x}_{pi})^{\boldsymbol{T}}, \boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \cdots, \boldsymbol{\alpha}_p)^T$ merupakan vektor koefisien regresi berukuran $p \times 1, p$ merupakan banyaknya variabel prediktor, $b_i$ merupakan konstanta positif yang diketahui, $v_i$ adalah pengaruh acak spesifik yang diasumsikan memiliki distribusi normal $v_i \sim N(0, \sigma_v^2)$.

Parameter $\boldsymbol{\theta}_i$, dapat diketahui dengan mengasumsikan bahwa model penduga langsung $\hat{\theta}_i$ telah tersedia pada persamaan 2

$$\widehat{\boldsymbol{\theta}}_i = \boldsymbol{\theta}_i + \boldsymbol{e}_i \,, \quad i = 1, \cdots, m \tag{2.2}$$

Dengan $\boldsymbol{e}_i \sim N(\boldsymbol{0}, \boldsymbol{\sigma}_\varepsilon^2)$

Rao (2003) menjelaskan bahwa model SAE untuk level area terdiri dari dua komponen model yaitu komponen model pendugaan langsung dan pendugaan tidak langsung. Kombinasi model pendugaan langsung (2.1) dan tak langsung (2.2) dikenal dengan *generalized linear mixed model* sebagai berikut:

$$\boldsymbol{\theta}_i = \boldsymbol{x}_i^T \boldsymbol{\alpha} + \boldsymbol{b}_i \boldsymbol{v}_i + \boldsymbol{e}_i \tag{2.3}$$

Model pada persamaan (2.3) dikenal sebagai model Fay-Heriot, dimana keragaman peubah respon didalam area kecil diasumsikan dapat diterangkan oleh hubungan antara peubah respon dengan informasi tambahan yang disebut model pengaruh tetap (*fixed effect*) yaitu $\boldsymbol{\alpha}$**.** Selain itu terdapat komponen keragaman spesifik area kecil yang tidak dapat diterangkan oleh informasi tambahan yang disebut dengan komponen pengaruh acak area kecil *(random effect*) yaitu $\boldsymbol{v_i}$**.**

### 2.1.2 Model Berbasis Unit

Model berbasis unit merupakan suatu model dimana data-data pendukung yang tersedia bersesuaian secara individu dengan data respon, misal $\boldsymbol{x_{ij}} = (x_{ij1}, \cdots, x_{ijp})^T$ dimana $\boldsymbol{x_{ij}}$ merupakan data penyerta untuk unit tertentu telah tersedia untuk masing-masing elemen populasi $j$ di setiap small area. Selanjutnya variabel $y_{ij}$, diasumsikan berhubungan dengan $\boldsymbol{x}_{ij}$ melalui model

$$\boldsymbol{y}_{ij} = \boldsymbol{x}_{ij}^{\boldsymbol{T}}\boldsymbol{\alpha} + \boldsymbol{v_i} + \boldsymbol{e}_{ij} \qquad (2.4)$$

Dengan $j = 1, \dots, n$ , $i = 1, \dots, m, v_i \sim N\left(0, \sigma_{e_i}^2\right)$.

Dimana $\boldsymbol{x_{ij}} = (x_{ij1}, \dots, x_{ijp})^T$ yang merupakan data penyerta unit tertentu, $p$ adalah variabel prediktor, $j$ adalah anggota rumah tangga/individu pada area ke-$i$, dan $v_i$ = pengaruh acak area yang disumsikan merupakan variabel yang bersifat *iid*

$e_{ia} = k_{ia} \times \tilde{e}_{ia}$

Dimana

$k_{ia}$: konstanta

$\tilde{e}_{ia}$ : variabel acak yang bersifat *iid* dan bebas terhadap $\boldsymbol{v_i}$**,** dimana $E_m(\tilde{e}_{ia}) = 0$ dan $V_{\varepsilon m}(\tilde{e}_{ia}) = \sigma_\varepsilon^2$

$v_i$ dan $e_{ia}$ seringkali diasumsikan memiliki distribusi peluang normal

Perbedaan mendasar pada kedua model tersebut yaitu pada penggunaan data pendukung yang tersedia. Pada model SAE level area, data pendukung yang tersedia hanya untuk level area tertentu. Model ini menghubungkan estimator langsung dengan variabel penyerta dari domain lain untuk setiap area. Sedangkan model level unit mengasumsikan bahwa variabel penyerta yang tersedia bersesuaian secara individu dengan variabel respon.

Penelitian ini mengembangkan model berbasis area, yakni model Fay-Herriot dengan pertimbangan ketersediaan data pada level unit hanya tersedia pada tahun-tahun pelaksanaan sensus penduduk sehingga sulit untuk melakukan estimasi pada tahun-tahun lainnya.

## *2.2* Model Regresi *Spline*

Regresi nonparametrik merupakan suatu metode statistika yang digunakan untuk mengetahui hubungan antara variabel respon dan prediktor yang tidak diketahui bentuk fungsinya, hanya diasumsikan fungsi *smooth* (mulus) dalam arti termuat dalam suatu ruang fungsi tertentu, sehingga regresi nonparametrik memiliki fleksibilitas yang tinggi (Eubank, 1988). Model regresi nonparametrik secara umum dapat disajikan sebagai berikut:

$$y_i = m(x_i) + e_i \ , i = 1, 2, \dots, n \tag{2.5}$$

Dengan $y_i$ adalah variabel respon, fungsi $m(x_i)$ adalah fungsi yang smooth dimana tidak diketahui bentuknya. Variabel $x_i$ sebagai variabel prediktor dengan $e_i \sim N(0, \sigma^2)$. Agar dapat menangani struktur $m(x_i)$ yang tidak linear, didefinisikan $K$ buah knot $k_1, \cdots, k_k$ dan dengan mengambil basis fungsi *polynomial* terputus diperoleh model berikut:

$$m(x_i) = \beta_0 + \beta_i x_i + \cdots + \beta_p x_i^p + \textstyle\sum_{j=1}^{k} \gamma_j (x_i - k_j)_+^p,$$

Dengan $p$ adalah derajat spline, $(x_i - k_j)_+ = maks\ \{0, (x_i - k_j)\}, k_j$ dimana $j = 1, \dots, K$ merupakan himpunan titik knot. Dengan menetapkan $\boldsymbol{\beta} = (\beta_0, \dots, \beta_p)^T$ adalah vektor koefisien parametrik, $\boldsymbol{\gamma}$=$(\gamma_1, \dots, \gamma_k)^T$ adalah vektor koefisien spline, $= \begin{bmatrix} 1 & x_i & \cdots & x_i^p \end{bmatrix}_{1 \le i \le n}, Z = [((x_i - k_1) \quad \cdots \quad (x_i - k_K)_+^p]]_{1 \le i \le n}$ ,

Dengan $(x_i - k_j)_+^p \begin{cases} = (x_i - k_j)_+^p & \text{untuk } x_i \ge k_j \\ = 0 & \text{untuk } x_i < k_j \end{cases}$

Sehingga model (2.5) dapat ditulis sebagai berikut:

$$y_i = \beta_0 + \beta_i x_i + \cdots + \beta_p x_i^p + \textstyle\sum_{j=1}^{k} \gamma_j (x_i - k_j)_+^p + e_i$$

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Z\gamma} + \boldsymbol{e} \tag{2.6}$$

Dimana **Y**=$(y_1, \ldots, y_k)^T$

Model (7) disebut sebagai *regresi spline smoothing* (Opsomer, 2004). Dari bentuk matematis fungsi spline pada model tersebut menunjukkan bahwa spline merupakan model polinomial terputus, tetapi masih bersifat kontinu pada knot-knotnya.

Knot dapat diartikan sebagai suatu titik fokus dalam fungsi spline sedemikian sehingga kurva yang dibentuk tersegmen pada titik tersebut. Titik knot merupakan salah satu hal yang sangat penting dalam pendekatan spline. Strategi yang digunakan untuk memilih dan menentukkan lokasi knot dengan tepat sangat dibutuhkan agar diperoleh model spline yang optimal. Jika jumlah knot terlampau banyak maka model yang dihasilkan akan *overfitting*.

Salah satu metode pemilihan titik knot optimal adalah dengan menggunakan *Generalized Cross Validation* (GCV).

Definisi GCV dapat ditulis sebagai berikut:

GCV (($\boldsymbol{K}$)=$\frac{MSE(K)}{[n^{-1}trace(I-A(K))]^2}$

Dimana $MSE(\boldsymbol{K}) = n^{-1}y^t(\boldsymbol{I} - \boldsymbol{A}(\boldsymbol{K}))^T\big(\boldsymbol{I} - \boldsymbol{A}(\boldsymbol{K})\big)\boldsymbol{y}, \boldsymbol{K} = \big(K_1, K_{2,} \cdots, K_N\big)$adalah titik knot dan matriks $\boldsymbol{A}(\boldsymbol{K})$ diperoleh dari persamaan $\hat{y} = A(\boldsymbol{K})\boldsymbol{y}$**.**

**2.3 Regresi *Penalized Spline***

Regresi *penalized spline* yaitu regresi yang diperoleh berdasarkan kuadrat terkecil (least square) dengan penalty kekasaran. *Penalized spline* mempunyai banyak kesamaan dengan smoothing spline, tetapi jenis penalty yang digunakan pada *penalized spline* lebih umum dibandingkan pada *smoothing spline* (Ruppert, 2003).

Menurut Hall dan Opsomer (2005), regresi *penalized spline* merupakan suatu pendekatan *smoothing* yang popular karena kesederhanaannya dan fleksibilitasnya. Pemodelan *penalized spline* memberikan pemilihan knot yang fleksibel. Salah satu alternatif untuk mengoptimalkan fit model terhadap data adalah dengan menambahkan penalty pada parameter spline. Dengan cara ini dapat dipilih jumlah knot yang banyak dan mencegah *overfitting* dengan menempatkan kendala *(constraint)*.

Terdapat dua komponen penting dalam mengestimasi *penalized spline*, yang pertama adalah pemilihan karakter smoothing, sementara yang kedua adalah pemilihan jumlah knot dan lokasinya (Yao dan Lee, 2008). Pada persamaan (2.6) dapat dinyatakan ke dalam bentuk matriks yaitu

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Z\gamma} + \boldsymbol{e}$$

Dimana

$$\mathbf{y}=\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \boldsymbol{X} = \begin{bmatrix} 1 & x_1 & \dots & x_k^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_1 & x_n^p \end{bmatrix}, \boldsymbol{Z} = \begin{bmatrix} (x_1 - k_1)_+^p & \dots & (x_1 - k_K)_+^p \\ \vdots & \ddots & \vdots \\ (x_n - k_1)_+^p & \cdots & (x_1 - k_K)_+^p \end{bmatrix},$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_p \end{bmatrix}, \ \boldsymbol{\gamma} = \begin{bmatrix} \gamma_1 \\ \vdots \\ \gamma_k \end{bmatrix}, \text{dan } \boldsymbol{e} = \begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix}$$

Estimator *penalized spline* diperoleh dengan meminimumkan fungsi *penalized least square* (PLS) sebagai berikut:

$$L = ||\boldsymbol{y} - \boldsymbol{X}\beta - \boldsymbol{Z}\gamma||^2 + \lambda\boldsymbol{\gamma}^T\boldsymbol{\gamma} \quad (2.7)$$

Dengan memisalkan $\boldsymbol{C} = [\boldsymbol{X}, \boldsymbol{Z}]$ dan $\ddot{\theta} = \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\gamma} \end{bmatrix}$ sehingga persamaan 9 dapat ditulis sebagai berikut:

$$L = ||\boldsymbol{y} - \boldsymbol{C}\ddot{\theta}||^2 + \lambda\ddot{\theta}^T\boldsymbol{D}\ddot{\theta} \quad (2.8)$$

Dimana diketahui **D** merupakan matrik penalty

$$\boldsymbol{D} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{0}_{(P+1)\times 2} & \mathbf{0}_{(P+1)\times K} \\ \mathbf{0}_{K\times(P+1)} & \boldsymbol{I}_{K\times K} \end{bmatrix}$$

Dengan parameter $\lambda$ parameter smoothing, dimana $\lambda \geq 0$. Suku pertama pada persamaan 10 adalah jumlah kuadrat error dan suku keduanya adalah penalty kekasaran. Menurut Djuraidah, et al (2006) Estimator *penalized spline* yang diperoleh adalah

$$\hat{\ddot{\Theta}} = (\boldsymbol{C}^T\boldsymbol{C} + \lambda\boldsymbol{D})^{-1}\boldsymbol{C}^T\boldsymbol{y} \tag{2.9}$$

Dengan demikian didapatkan $\hat{\boldsymbol{y}} = \boldsymbol{C}\hat{\ddot{\Theta}}$

$$\hat{\boldsymbol{y}} = \boldsymbol{C}(\boldsymbol{C^T}\boldsymbol{C} + \boldsymbol{\lambda}\boldsymbol{D})^{-\boldsymbol{1}}\boldsymbol{C^T}\boldsymbol{y} \tag{2.10}$$

Berdasarkan uraian di atas, nilai $\hat{\theta}$ bergantung pada parameter smoothing $\lambda$. Jika nilai $\lambda$ besar akan menghasilkan bentuk kurva regresi yang sangat halus. Sebaliknya, jika nilai $\lambda$ kecil akan memberikan bentuk kurva regresi yang sangat kasar. Akibatnya pemilihan parameter smoothing optimal perlu dilakukan. Dengan menggunakan generalized cross-validation (GCV) yang didefinisikan sebagai berikut:

$$GCV(\lambda) = \frac{n^{-1}RSS(\lambda)}{(1-n^{-1}df_\lambda)^2} = \frac{MSE(\lambda)}{(n^{-1}tr(\boldsymbol{I}-\boldsymbol{S_\lambda}))^2} \tag{2.11}$$

Dimana $RSS(\lambda) = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$, $df_\lambda = tr(\boldsymbol{S}_\lambda)$

$\boldsymbol{S}_\lambda = \boldsymbol{C}(\boldsymbol{C^T}\boldsymbol{C} + \boldsymbol{\lambda}\boldsymbol{D})^{-\boldsymbol{1}}\boldsymbol{C^T}$ yang disebut dengan matriks *smoothing* (Ruppert, et al., 2003; Griggs,2013)

Pada penelitian ini untuk melakukan penetuan jumlah titik knot dapat dilakukan dengan metode *fixed selection method*. Tujuan utama untuk semua metode pemilihan knot Ҡ adalah untuk memastikan bahwa Ҡ cukup besar agar lebih fleksibel ketika mengontrol kemulusan kurva yang diestimasi dengan smoothing parameter. Tujuan lainnya adalah memilih Ҡ yang tidak terlalu besar agar waktu perhitungan yang dibutuhkan tidak terlalu lama atau MSE yang lebih besar dari seharusnya. Rumus *fixed selection method* didefinisikan sebagai berikut:

$$Ҡ = \min(\tfrac{1}{4} \times \text{banyaknya } x_i \text{yang } unique, 35) \tag{2.12}$$

Persamaan diatas merupakan metode yang umumnya digunakan untuk pemilihan jumlah knot dan penentuan lokasi knot yang optimum ditentukkan melalui kuantil ke-$K_k$ dari $x_i$ yang *unique*, dengan rumus sebagai berikut (Ruppert, et al., 2003)

$$K_k = \left(\frac{k+1}{K+2}\right), k = 1,2, \dots, K \tag{2.13}$$

## 2.4 Pendugaan Area Kecil dengan Pendekatan Semiparametrik *Penalized Spline*

Pendugaan area kecil (SAE) adalah pendekatan yang digunakan untuk mengungkapkan hubungan antara variabel interest dengan variabel pendukung sebagai model linear dengan tambahan pengaruh acak area kecil. Dimisalkan $\theta$ merupakan vektor dari parameter *small area* yang berukuran $m\times1$ dan diasumsikan vektor tersebut merupakan estimator langsung $\hat{\theta}$. Jika dinyatakan $m\times q$ adalah matriks dari variabel penyerta dari level area $\boldsymbol{x}_i = (\boldsymbol{x}_{1i,}\boldsymbol{x}_{2i,\dots}\boldsymbol{x}_{pi})^{\boldsymbol{T}}$ sehingga model SAE berbasis area dapat ditulis seperti persamaan (2.3) adalah sebagai berikut:

$$\boldsymbol{\theta}_i = \boldsymbol{x}_i^T\boldsymbol{\alpha} + \boldsymbol{b}_i\boldsymbol{v}_i + \boldsymbol{e}_i \; ; \quad i = 1,2, \dots, m \; ; \; v_i \sim N(0, \sigma_v^2)$$

dimana $b_i$ merupakan konstanta positif yang diketahui, $v_i$ adalah pengaruh acak spesifik yang diasumsikan memiliki distribusi normal $v_i \sim N(0, \sigma_v^2)$. Menurut Giusti et al (2012), model SAE berbasis area ini menghasilkan estimasi *small area* yang terpercaya dengan mengkombinasikan model SAE dan dan model regresi yang meminjam kekuatan dari domain lain, ketika asumsi ini tidak terpenuhi model SAE level area menyebabkan estimator bias dari parameter daerah kecil. Spesifikasi semiparametrik dari model SAE yang memungkinkan yaitu adanya hubungan nonlinear antara $\hat{\boldsymbol{\theta}}$ dan variabel penyerta $\boldsymbol{X}$, dapat diperoleh dengan menggunakan pendekatan *penalized spline*.

Seperti pada persamaan (2.5), model semiparametrik dengan satu respon $x_1$ dapat ditulis $\widetilde{m}(x_1)$ dimana fungsi dari $\widetilde{m}(.)$ tidak diketahui akan tetapi diasumsikan cukup baik sehingga diberikan fungsi spline adalah sebagai berikut:

$$m(x_1) = \beta_0 + \beta_i x_1 + \cdots + \beta_p x_1^p + \sum_{j=1}^{k} \gamma_j (x_i - k_j)_+^p,$$

Dengan $p$ adalah derajat spline, $(x_i - k_j)_+ = maks\,\{0, (x_i - k_j)\}, k_j$ dimana $j = 1, \dots, K$ merupakan himpunan titik knot. Dengan menetapkan $\boldsymbol{\beta} = (\beta_0, \dots, \beta_p)^T$ adalah $(p+1)$ vektor koefisien fungsi polinomial, $\boldsymbol{\gamma}$=$(\gamma_1, \dots, \gamma_k)^T$ adalah vektor koefisien spline,

Dengan $(x_i - k_j)_+^p \begin{cases} = (x_i - k_j)_+^p & \text{untuk } x_i \geq k_j \\ = 0 & \text{untuk } x_i < k_j \end{cases}$

Menurut opsomer, et al (2008) model tersebut diindikasikan akan *overparameterized* sehingga akan menyebabkan *overfitiing* untuk menghindari hal tersebut ditambahkan penalty pada parameter spline dengan meminimumkan fungsi *penalized least square* sehingga didapatkan hasil sesuai dengan persamaan (2.10).

Pada penelitian ini pendekatan SAE dengan menggunakan *penalized spline* sebagai efek random menghasilkan

$$y_i = m(\boldsymbol{X}_i; \boldsymbol{\beta}) + \boldsymbol{\varepsilon}_i$$

$$= \boldsymbol{X}_i * \boldsymbol{\beta} + \boldsymbol{\varepsilon}_i$$

$$= \boldsymbol{X}_i \boldsymbol{\beta} + \boldsymbol{Z}_i \boldsymbol{\gamma} + \boldsymbol{\varepsilon}_i \qquad (2.14)$$

Dimana: $\boldsymbol{X}_i \boldsymbol{\beta} = \beta_0 + \beta_1 X_i + \cdots + \beta_p X_i^p$ (komponen parametrik yang merupakan *fixed component*); $\boldsymbol{Z}_i \boldsymbol{\gamma} = Z_{1i}\gamma_1 + \cdots + Z_{ki}\gamma_k$=$((X_i - K_1)_+^P \beta_{pi1} + \cdots + (X_i - K_1)_+^P \beta_{piK}$ (deviasi dari komponen parametrik dengan random efek); dan $\gamma = (\gamma_1, \dots, \gamma_k)$ yang berasumsi mean 0 dan varians $\sigma_\gamma^2$. Model *penalized spline* merupakan model *random effect* yang dapat dikombinasikan dengan model SAE berbasis area agar mendapatkan estimasi pendugaan area kecil secara semiparametrik berdasarkan *linear mixed model*. Dari persamaan (2.3) dan persamaan (2.14) didapatkan model semiparametrik Fay-Herriot dapat ditulis sebagai berikut:

$$\hat{\theta} = \begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{X_1} \end{bmatrix} [\alpha, \beta] + \boldsymbol{Z}\gamma + \boldsymbol{b}v + e \qquad (2.15)$$

Menurut Giusti et al (2012), jika terdapat variabel lain yang perlu disertakan dalam model, variabel tersebut dapat ditambahkan kedalam $X$ sebagai matriks efek tetap. Opsomer (2004) menggunakan *penalized spline* untuk mengestimasi area kecil dan menambahkan pengaruh acak kecil pada model sehingga didapatkan persamaan:

$$\hat{\theta} = \boldsymbol{X}\beta + \boldsymbol{Z}\gamma + \boldsymbol{b}v + e \tag{2.16}$$

Persamaan diatas terdiri dari fungsi spline yang merupakan fungsi semiparametrik $\boldsymbol{X}\beta + \boldsymbol{Z}\gamma$ dan pengaruh acak area kecil $(\boldsymbol{b}v)$. Nilai estimasi pada $\tilde{\beta}$ untuk regresi semiparametrik *penalized spline* untuk penduga area kecil dengan menggunakan (MLE) sehingga didapatkan

$$\hat{\beta} = (\boldsymbol{X^T V^{-1} X})^{-1} \boldsymbol{X^T V^{-1} Y}$$

Jika komponen varians tidak diketahui, maka setelah estimator $\beta$ dan prediktor $\gamma$ diperoleh, Estimasi komponen varians berdasarkan ML bias, maka digunakan metode REML (*Restricted Maximum Likelihood*).

## 2.5 *Best Linear Unbiased Prediction* (BLUP) dan *Empirical Best Linear Unbiased Prediction* (EBLUP)

Model *small area* terbagi menjadi model area level dan model unit level. Metode BLUP dan EBLUP salah satu metode yang digunakan untuk meminimumkan MSE. Pada metode BLUP, variansi pengaruh acak diasumsikan telah diketahui. Sedangkan pada metode EBLUP nilai variansi pengaruh acak small area tidak diketahui sehingga harus ditaksir dengan menggunakan metode *Maximum Likelihood* (ML)

Misalkan data memenuhi model linear campuran berikut:

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{Z\gamma} + \boldsymbol{e} \tag{2.17}$$

Dimana

$\boldsymbol{y}$ adalah vektor data observasi berukuran $n \times 1$

$\boldsymbol{X}$ dan $\boldsymbol{Z}$ adalah matriks berukuran $n \times p$ dan $n \times h$ yang diketahui

$\boldsymbol{\gamma}$ dan $\boldsymbol{e}$ adalah berdistribusi saling bebas dengan rataan 0 dan ragam $\boldsymbol{G}$ dan $\boldsymbol{R}$ yang tergantung pada parameter $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_q)^T$, diasumsikan bahwa $\boldsymbol{\delta}$ adalah himpunan bagian dari ruang *Euclidean* sedemikian sehingga:

$$\boldsymbol{Var}(\boldsymbol{y}) = \boldsymbol{V} = \boldsymbol{V}(\boldsymbol{\delta}) = \boldsymbol{R} + \boldsymbol{ZGZ^T}$$

Adalah non singular untuk semua $\delta$ yang terdapat dalam himpunan bagian tersebut, dimana $Var(\boldsymbol{y})$ adalah matriks varians covarians dari $\boldsymbol{y}$.

Parameter yang akan diduga merupakan kombinasi linear $\boldsymbol{\mu} = \boldsymbol{1_l^T}\boldsymbol{\beta}$+$\boldsymbol{m^T}\boldsymbol{v}$ (Rao,2003). Vektor $\boldsymbol{1}$ dan $\boldsymbol{m}$ adalah konstan. Penduga linear dari $\mu$ adalah $\hat{\boldsymbol{\mu}} = \boldsymbol{a^T}\boldsymbol{\beta} + \boldsymbol{b}$ untuk $\boldsymbol{a}$ dan $\boldsymbol{b}$ diketahui. Sehingga penduga tak bias $\boldsymbol{\mu}$

$E(\hat{\mu}) = E(\mu)$

$E$ adalah ekpektasi, MSE $\hat{\mu}$ didefinisikan sebagai $MSE(\hat{\mu}) = E(\hat{\mu} - \mu)^2$

Jika $\hat{\mu}$ adalah penduga tak bias dari $\mu$, maka $MSE(\hat{\mu}) = E(\hat{\mu} - \mu)^2 = Var\ (\hat{\mu} - \mu)^2$

Estimator BLUP $\mu$ yang $\delta$ diketahui adalah sebagai berikut:

$$\tilde{\boldsymbol{\mu}}^{\boldsymbol{H}} = \boldsymbol{t}(\boldsymbol{\delta}, \boldsymbol{y}) = \boldsymbol{I^T}\tilde{\boldsymbol{\beta}} + \boldsymbol{m^T}\tilde{\boldsymbol{v}} = \boldsymbol{I^T}\tilde{\boldsymbol{\beta}} + \boldsymbol{m^T GZ^T V^{-1}}(\boldsymbol{y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}}) \tag{2.18}$$

Dimana

$$\tilde{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}(\boldsymbol{\delta}) = (\boldsymbol{X^T V^{-1} X})^{-1}\boldsymbol{X^T V^{-1} y} \tag{2.19}$$

merupakan *best linear unbiased* estimator (BLUE) dari $\beta$ dan

$$\tilde{\boldsymbol{v}} = \tilde{\boldsymbol{v}}(\boldsymbol{\delta}) = \boldsymbol{GZ^T V^{-1}}(\boldsymbol{y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}}) \tag{2.20}$$

Keterangan $H$ pada $\tilde{\mu}$ adalah Henderson yang mengusulkan persamaan (2.18)

Penduga BLUP tergantung pada ragam $\delta$ yang biasanya tidak diketahui. Jika $\delta$ diduga dengan $\tilde{\delta} = \tilde{\delta}(y)$, maka akan diperoleh *empirical best linear unbiased prediction* (EBLUP) yang tetap merupakan penduga tak bias bagi $\mu$. Penduga $\delta$ diperoleh melalui metode ML atau REML (Rumiati, 2012).

## 2.6 Pendugaan MSE dengan menggunakan Metode *Jackknife* dan Pendugaan MSE tidak langsung

Menurut Baillo dan Molina (2009), tujuan dari prosedur dan teknik yang digunakan dalam SAE adalah untuk memperoleh estimasi dengan tingkat presisi yang tinggi pada area kecil tersebut. Tingkat presisi estimator ini dapat digambarkan oleh *Mean Square Error* (MSE). Penerapan *jackknife* pada SAE dilakukan untuk mengkoreksi pendugaan MSE.

Fay dan Herriot (1979) mengembangkan model $y_i = x_i^T\beta + v_i + e_i$ sebagai dasar dalam pengembangan SAE. Untuk selanjutnya diasumsikan bahwa $\beta$ dan $\sigma_v^2$ tidak diketahui, akan tetapi $\sigma_{ei}^2$ diketahui, dengan $\beta_i = \sigma_{ei}^2/(\sigma_v^2 + \sigma_{ei}^2)$ maka

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

$$= Var\left(\hat{\theta}_i^{EBLUP}\right) + (bias(\hat{\theta}_i^{EBLUP}))^2$$

Persamaan tersebut dapat diuraikan menjadi

$$MSE\left(\hat{\theta}_i^{EBLUP}\right) = MSE\left(\hat{\theta}_i^{BLUP}\right) + E(\hat{\theta}_i^{EBLUP} - \hat{\theta}_i^{BLUP})^2 \qquad (2.21)$$

Metode *jackknife* pertama kali diperkenalkan oleh tukey pada tahun 1958 dan kemudian berkembang sebagai suatu metode untuk mengoreksi bias pada suatu estimator. Dengan melakukan penghapusan terhadap observasi ke-$i$ untuk $i = 1,2,\dots,m$ dan kemudian dilakukan pendugaan parameter misal $\hat{\theta}_{(i)}$, maka penduga bias diduga dengan

$$bias\left(\hat{\theta}\right) = (m-1)\left[\hat{\theta}_{(.)} - \hat{\theta}\right]$$

dengan $\hat{\theta}_{(.)} = m^{-1}\sum_i^m \hat{\theta}_{(i)}$

Penduga *jackknife* diperoleh dari

$$\hat{\theta}_{jack} = \hat{\theta} - bias(\hat{\theta}) \text{ dan } v_{jack}\left(\hat{\theta}\right) = \frac{(n-1)}{n}\sum_i^m\left[\hat{\theta}_{(i)} - \hat{\theta}\right]^2$$

Penerapan *jackknife* pada SAE dilakukan untuk mengkoreksi pendugaan MSE akibat adanya pendugaan $\alpha$ dan $\sigma_v^2$. Persamaan (2.19) setara dengan $g_{1i}(\sigma_v^2) + (bias)^2$ jika $\sigma_v^2$ diduga.

Dengan $u$ adalah banyak replikasi *jackknife* dan $i$ adalah banyak data, maka prosedur *jackknife* $MSE\left(\hat{\theta}_i^{EBLUP}\right)$ pendugaan tidak langsung berdasarkan persamaan (2.21) adalah sebagai berikut:

1. $MSE\left(\hat{\theta}_i^{EBLUP}\right)$ didekati oleh:

   $MSE_i\left(\hat{\theta}_i\right) = h_{1i} + h_{2i}$
2. Menduga variansi $MSE\left(\hat{\theta}_i^{EBLUP}\right)$ dengan menghitung:

   $h_{1i} = g_{1i}(S_v^2) - \left(\frac{m-1}{m}\right)\sum_{u=1}^{m}\left[g_{1i}\left(S_{v(-u)}^2\right) - g_{1i}(S_v^2)\right]$

   Dimana $g_{1i}\left(S_{v(-u)}^2\right)$ diperoleh dengan menghapus pengamatan ke-$u$ pada himpunan data $g_{1i}(S_v^2)$ dan $u = 1,2,\dots,m$. Dengan:

   $S_v^2 = (m-1)^{-1}\sum_i (y_i - \bar{y})^2 - \sigma_e^2$

   $S_{v(-u)}^2 = (m-2)^{-1}\sum_{i(-u)}\left(y_i - \bar{y}_{(-u)}\right)^2 - \sigma_e^2$
3. Menduga $E\left(\hat{\theta}_i^{EBLUP} - \hat{\theta}_i^{BLUP}\right)^2$ dengan menghitung:

   $h_{2i} = \left(\frac{m-1}{m}\right)\sum_{u=1}^{m}\left[\left(\hat{\theta}_{i(-u)}\right) - (\hat{\theta}_i)\right]^2$

   Dimana $\left(\hat{\theta}_{i(-u)}\right)$ diperoleh dengan menghapus pengamatan ke-$u$ pada himpunan data $(\hat{\theta}_i)$

Untuk membandingkan hasil estimasi tidak langsung pada *small area estimation* dapat menggunakan nilai MSE dengan metode pendugaan langsung dengan resampling *jackknife.*

Nilai MSE pendugaan langsung dengan $u$ adalah banyak replikasi *jackknife* dan $i$ adalah banyak data, maka prosedur *jackknife* pendugaan langsung berdasarkan persamaan (2.21) adalah sebagai berikut:

1. $MSE$ pendugaan langsung didekati oleh:

   $MSE_i(y_i) = h_{1i} + h_{2i}$
2. Menduga variansi $MSE$ pendugaan langsung dengan menghitung:

$h_{1i} = g_{1i}(S_v^2) - \left(\frac{m-1}{m}\right) \sum_{u=1}^{m} \left[ g_{1i}\left(S_{v(-u)}^2\right) - g_{1i}(S_v^2) \right]$

Dimana $g_{1i}\left(S_{v(-u)}^2\right)$ diperoleh dengan menghapus pengamatan ke- $u$ pada himpunan data $g_{1i}(S_v^2)$ dan $u = 1,2, \dots, m$. Dengan:

$S_v^2 = (m-1)^{-1} \sum_i (y_i - \bar{y})^2 - \sigma_e^2$

$S_{v(-u)}^2 = (m-2)^{-1} \sum_{i(-u)} \left(y_i - \bar{y}_{(-u)}\right)^2 - \sigma_e^2$

3. Menduga nilai $h_{2i}$ dengan menghitung:

$h_{2i} = \left(\frac{m-1}{m}\right) \sum_{u=1}^{m} \left[ \left(y_{i(-u)}\right) - (y_i) \right]^2$

Dimana $\left(y_{i(-u)}\right)$ diperoleh dengan menghapus pengamatan ke-$u$ pada himpunan data $(y_i)$. Nilai RMSE diperoleh setelah mendapatkan nilai MSE melalui persamaan (2.23)

$$RRMSE\left(\hat{\theta}_i\right) = \frac{\sqrt{MSE\left(\hat{\theta}_i\right)}}{\hat{\theta}_i} \times 100\% \qquad (2.22)$$

## 2.7 Survei Sosial Ekonomi Nasional (SUSENAS)

Susenas merupakan survei yang dirancang untuk mengumpulkan data sosial kependudukan yang relative sangat luas. Susenas merupakan salah satu survei yang rutin dilakukan oleh Badan Pusat Statistik (BPS). Melalui Susenas dikumpulkan data yang berkaitan dengan kondisi sosial ekonomi masyarakat meliputi kondisi kesehatan, pendidikan, fertilitas, keluarga berencana, perumahan, dan kondisi sosial ekonomi lainnya. Adapun tujuan penyelenggaraan susenas antara lain:

1. Menyediakan data yang berkaitan dengan kondisi sosial ekonomi masyarakat meliputi kondisi kesehatan, pendidikan, fertilitas, keluarga berencana, perumahan dan kondisi sosial ekonomi lainnya. Data dan indikator dari Susenas telah dipergunakan secara luas dan dipandang sebagai salah satu bukti penting yang dapat berguna untuk perencanaan, monitoring dan evaluasi program pembangunan pemerintah.

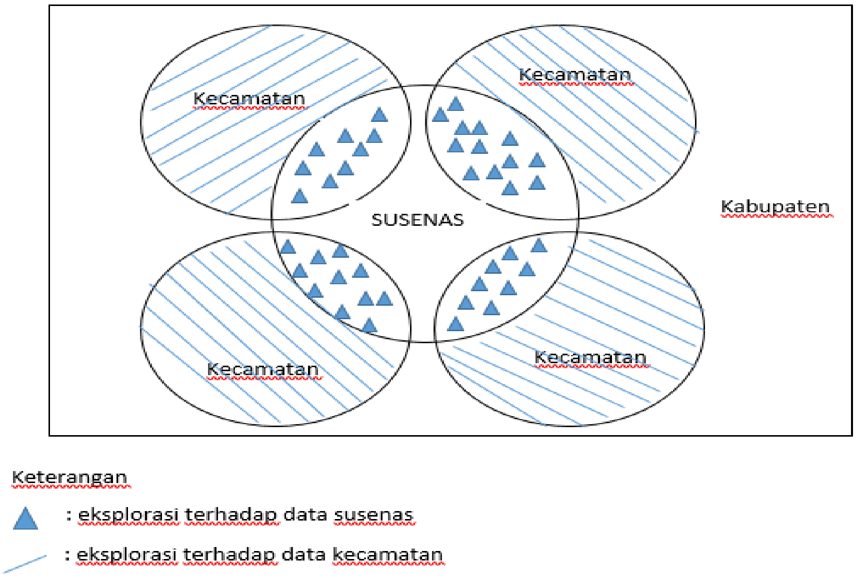
2. Memenuhi kebutuhan pemerintah, khususnya untuk penyediaan data tingkat kemiskinan dalam interval waktu yang lebih pendek (dari sebelumnya sekali setahun menjadi dua kali setahun atau lebih), maka mulai tahun 2011 BPS melakukan perubahan dalam penyelenggaraan Susenas. Perubahan penting dalam penyelenggaraan Susenas 2011, dan masih diteruskan sampai tahun 2014 yaitu hasil statistik yang dihasilkan susenas antara lain

1. Statistik/Indikator Kesejahteraan Rakyat (Kesra) Statistik/Indikator Kesra yang dapat disusun dari hasil pengumpulan data kor, antara lain adalah Angka Partisipasi Sekolah, Rata-rata Lama Sekolah, Angka Melek Huruf (bidang pendidikan), Angka Kesakitan (bidang kesehatan), Rata-rata Umur Perkawinan Pertama, Angka Partisipasi KB (bidang fertilitas), Rata-rata Luas Hunian Rumah perkapita, Persentase Penggunaan Air Bersih (bidang perumahan), data publikasi wanita dan pria, dan lain-lain.

2. Statistik Konsumsi dan Pengeluaran Statistik yang dapat disusun dari hasil pengumpulan data konsumsi, antara lain rata-rata pengeluaran penduduk yang dirinci menurut jenis makanan dan bukan makanan, rata-rata konsumsi penduduk yang dirinci menurut jenis makanan, rata-rata konsumsi kalori dan protein, angka gini ratio, dan jumlah/persentase penduduk miskin.

Pada penelitian ini, konsep susenas yang digunakan dalam *small area estimation* adalah sebagai berikut:

Konsep susenas yaitu representative dari kabupaten, dimana pengambilan sampelnya yaitu sejumlah rumah tangga yang sudah terdapat pada data blok sensus kemudian digunakan untuk melakukan survey dengan sejumlah data yang telah dieksplorasi sebelumnya. Dengan penjelasan seperti pada gambar 2.1




Gambar 2.1 Ilustrasi Susenas

Kemudian pada gambar 2.1 tersebut jika dikaitkan dengan penelitian yang menggunakan small area estimation untuk level area pada kabupaten Sleman,, yaitu desain sampling pada susenas hanya untuk menggambarkan satu kabupaten dan tidak menggambarkan level kecamatan, pada penelitian ini Kabupaten Sleman memiliki kecamatan sebanyak 17 kecamatan. Pengambilan sampel pada susenas tidak semua kecamatan tersurvei, ada juga kecamatan yang tidak tersurvei. Di Kabupaten Sleman ini sampel yang diambil pada susenas terdapat pada tabel 2.1.

Tabel 2.1.a Daftar Pengambilan Sampel pada Susenas Kabupaten Sleman

| No | Kecamatan | Jumlah RT | Jumlah Anggota RT |
|---|---|---|---|
| 1 | Berbah | 30 | 98 |
| 2 | Cangkringan | 30 | 100 |
| 3 | Depok | 168 | 344 |
| 4 | Gamping | 59 | 215 |
| 5 | Godean | 57 | 210 |
| 6 | Kalasan | 39 | 135 |
| 7 | Minggir | 26 | 89 |
| 8 | Mlati | 74 | 226 |
| 9 | Moyudan | 20 | 63 |
| 10 | Ngaglik | 78 | 240 |
| 11 | Ngemplak | 57 | 171 |

Tabel 2.1.b Daftar Pengambilan Sampel pada Susenas Kabupaten Sleman

| No | Kecamatan | Jumlah RT | Jumlah Anggota RT |
|---|---|---|---|
| 12 | Pakem | 20 | 87 |
| 13 | Prambanan | 28 | 91 |
| 14 | Seyegan | 30 | 117 |
| 15 | Sleman | 39 | 138 |
| 16 | Tempel | 50 | 170 |
| 17 | Turi | 10 | 45 |

Sumber : diolah dari SUSENAS 2014

Pada tabel 2.1 tersebut merupakan sampel yang tersurvei dan digunakan pada susenas, terdapat 815 Rumah Tangga dan 2.539 Anggota Rumah Tangga. Sampel yang diambil tidak dapat mewakili kabupaten yang akan diteliti, terdapat beberapa sampel pada susenas yang terambil hanya 10-26 rumah tangga, sampel ini tidak dapat mewakili untuk menggambarkan secara keseluruhan kecamatan tersebut. Jika peneliti menggunakan data susenas untuk mengestimasi kecamatan di Kabupaten Sleman akan mempunyai error yang besar dan hasil penduga indikator menjadi kurang dipercaya. Oleh karena itu SAE menjadi solusi untuk mengatasi masalah estimasi pada kecamatan di Kabupaten Sleman.

## 2.8 Pengeluaran Perkapita

Besarnya pendapatan perkapita dapat menggambarkan kesejahteraan suatu masyarakat. Namun data pendapatan yang akurat sulit diperoleh, sehingga dalam kegiatan susenas data ini didekati melalui pengeluaran perkapita (BPS, 2010).

Pengeluaran perkapita menunjukkan besarnya pengeluaran setiap anggota rumah tangga dalam kurun waktu satu bulan (BPS, 2008). Maksud dari rumah tangga sendiri adalah sekelompok orang yang mendiami sebagian atau seluruh bangunan fisik dan biasanya tinggal bersama serta makan dari satu dapur. Dimana pengeluaran perkapita dipengaruhi oleh pendapatan perkapitanya. Asumsi ini menjelaskan pada saat pendapatan seseorang semakin meningkat maka semakin tinggi pula pengeluarannya.

Berdasarkan pedoman pencacah modul konsumsi SUSENAS 2002, dalam sensus pengeluaran perkapita merupakan pengeluaran untuk rumah tangga/anggota rumah tangga saja, tidak termasuk pengeluaran untuk keperluan usaha rumah tangga, atau yang diberikan kepada orang lain. Untuk konsumsi makanan, baik banyaknya (kuantitas) maupun nilainya yang dicatat adalah betul-betul telah dikonsumsi selama referensi waktu survey (*consumption approach*), sedangkan untuk bukan makanan konsep yang dipakai pada umumnya adalah konsep pembelian (*delivery approach*), yaitu dicatat sebagai pengeluaran pada waktu barang tersebut dibeli/diperoleh, asalkan tujuannya untuk kebutuhan rumah tangga.

# BAB 3
# METODOLOGI PENELITIAN

## 3.1 Sumber Data

Sumber data yang digunakan dalam penulisan ini menggunakan mikro data Susenas, Potensi Desa (Podes), dan data sekunder yang berasal dari publikasi-publikasi yang dihasilkan oleh BPS. Data yang digunakan adalah data tahun 2014 di 17 kecamatan di Kabupaten Sleman.

## 3.2 Variabel Penelitian

Variabel penelitian yang digunakan dalam penelitian ini terdiri respon (Y), prediktor (X) sebanyak 5 variabel. Variabel secara rinci disajikan pada tabel 3.1

**Tabel 3.1 Variabel Penelitian**

| Variabel | Nama Variabel | Tipe Variabel |
|---|---|---|
| $Y$ | Pengeluaran per kapita rumah tangga sebulan | Kontinu |
| $X_1$ | Keluarga petani | Kontinu |
| $X_2$ | Keluarga pengguna listrik pln | Kontinu |
| $X_3$ | Jumlah Penduduk yang sedang sekolah | Kontinu |
| $X_4$ | Rata-rata anggota keluarga | Kontinu |
| $X_5$ | Kepadatan penduduk | Kontinu |

Adapun definisi dari variabel-variabel penelitian diatas menurut definisi istilah yang diterbitkan oleh BPS dan beberapa literatur adalah sebagai berikut :

1. Pengeluaran per kapita rumah tangga sebulan (Y), adalah total pengeluaran rumah tangga sebulan dibagi dengan banyaknya anggota rumah tangga (ART). Pengeluaran didefinisikan pengeluaran untuk kebutuhan rumah tangga/anggota rumah tangga saja, tidak termasuk konsumsi/pengeluaran untuk keperluan usaha rumah tangga, atau yang diberikan kepada pihak orang lain. Pada penelitian ini data pengeluaran per kapita di jumlahkan untuk berdasarkan kecamatan, data berasal dari susenas Sleman 2014.

2. Keluarga pertanian ($X_1$), yaitu keluarga yang sekurang-kurangnya terdapat satu anggota keluarga melakukan kegiatan pertanian per kecamatan di Kabupaten Sleman, data berasal dari Sleman dalam angka 2015.
3. Keluarga pengguna listrik PLN ($X_2$), yaitu banyaknya keluarga yang menggunakan listrik PLN per kecamatan di Kabupaten Sleman, dengan penjumlahan setiap keluarga pengguna listrik per kecamatan di Kabupaten Sleman, data berasal dari podes 2014.
4. Jumlah penduduk yang sedang sekolah ($X_3$), yaitu jumlah penduduk yang bersekolah di sekolah formal, mulai dari pendidikan dasar sampai dengan pendidikan tinggi selama seminggu per kecamatan di Kabupaten Sleman, data berasal dari Sleman dalam Angka 2015.
5. Rata-rata anggota keluarga ($X_4$), yaitu keluarga adalah unit terkecil dari masyarakat yang terdiri atas kepala keluarga dan beberapa orang yang terkumpul dan tinggal di suatu tempat di bawah suatu atap dalam keadaan saling ketergantungan per kecamatan di Kabupaten Sleman, data diambil dari Sleman dalam Angka 2015.
6. Kepadatan Penduduk ($X_5$), yaitu luas wilayah kabupaten dibagi dengan jumlah penduduk di wilayah tersebut per kecamatan di Kabupaten Sleman, data berasal dari Sleman dalam angka 2015.

## 3.3 Langkah Analisis Model Pendugaan Area Kecil Menggunakan Semiparametrik Penalized Spline

Sebelum melakukan tahap penelitian, terlebih dahulu dilakukan tahap *Pre-Processing* data yang akan diolah sebagai berikut:

1. Menyiapkan data pengeluaran per kapita kabupaten Sleman dari data SUSENAS 2014
2. Menyiapkan data untuk variabel prediktor ($X_1$, $X_3$, $X_4$,, dan $X_5$) per kecamatan di Kabupaten Sleman yang berasal dari Sleman dalam Angka 2015
3. Menyiapkan data untuk variabel prediktor ($X_2$) per kecamatan di Kabupaten Sleman dari data PODES 2014
4. Menggabungkan data pada tahap ke-2 dan tahap ke-3 menjadi satu set data

Selanjutnya metode dan tahapan penelitian yang akan dilakukan untuk mencapai tujuan penelitian adalah sebagai berikut:

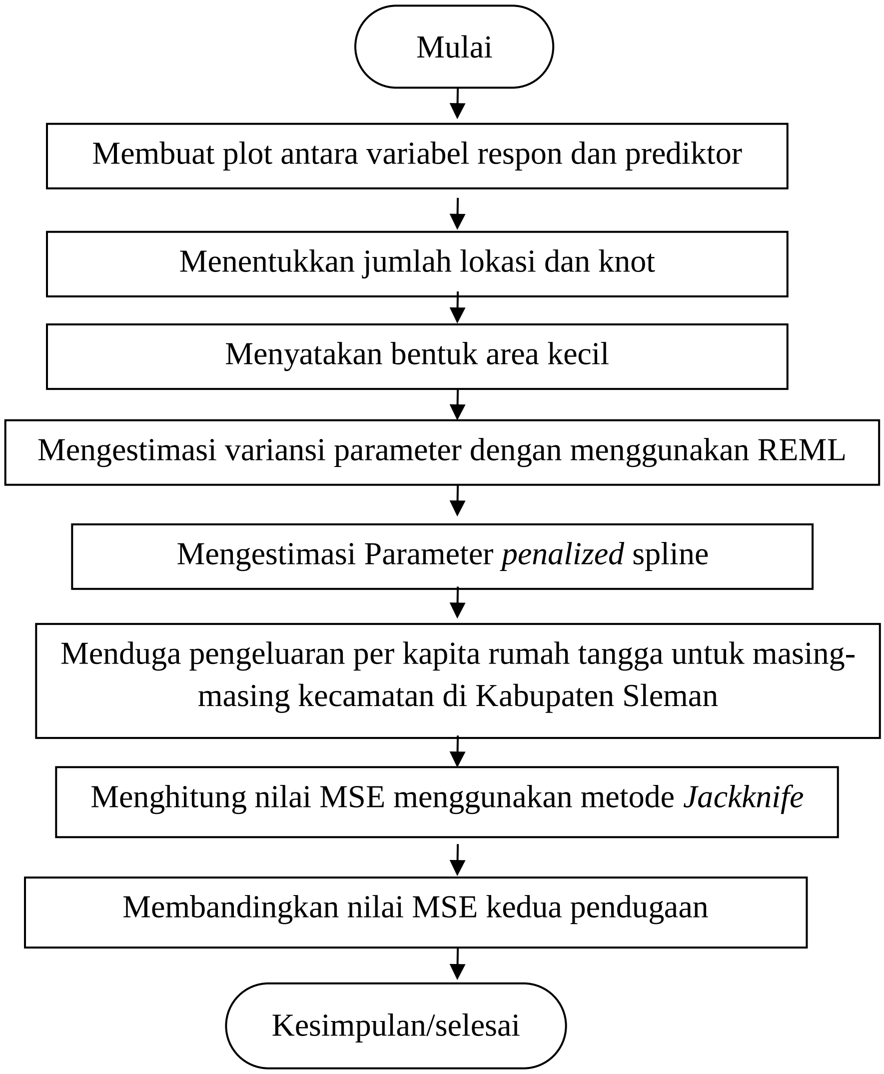
a) Eksplorasi data variabel penelitian;
   1. Eksplorasi data dan membuat statistik deskriptif untuk variabel respon dan variabel penyerta/prediktor
   2. Membuat plot data antara variabel respon dengan semua variabel prediktor secara parsial kemudian hasil tersebut ditentukkan komponen parametrik dan variabel komponen nonparametrik
   3. Menentukkan variabel prediktor yang menggunakan kurva parametrik linier dan kurva nonparametrik spline

b) Menduga pengeluaran per kapita dengan secara tidak langsung dengan menggunakan metode Semiparametrik *Penalized Spline* dilakukan sebagai berikut:
   1. Proses pembentukkan model menggunakan metode generalized additive model $y = f(x_1, x_2, x_3, x_4) + f(x_5) + u$, dimana mencari nilai parametrik untuk $f(parametrik)$, nilai dugaan parametrik digunakan untuk tahapan pencarian nilai nonparametrik *penalized* spline untuk $f(x_5)$
   2. Memodelkan komponen nonparametrik dengan pendekatan *penalized* spline, kemudian menentukkan jumlah knot yang digunakan dengan menggunakan *fixed selection method*
   3. Menghitung nilai GCV untuk masing-masing jumlah knot
   4. Memilih jumlah knot optimum berdasarkan nilai GCV minimum
   5. Menghitung nilai lambda (λ) menggunakan kriteria GCV
   6. Menyatakan bentuk area kecil dengan mengikuti suatu fungsi *Penalized* spline berbasis polynomial terputus $\hat{\theta} = \boldsymbol{X}\beta + \boldsymbol{Z}\gamma + \boldsymbol{b}v$
   7. Mengestimasi $\beta$, $\gamma$, dan $v$
   8. Menduga pengeluaran perkapita rata-rata untuk masing-masing kecamatan di Kabupaten Sleman

c) Menghitung MSE hasil penduga pengeluaran perkapita untuk masing-masing kecamatan dengan pendekatan *jackknife* dan juga menggunakan MSE langsung

## 3.4 Diagram Alir Penelitian

Diagram alir dalam penelitian ini diperlihatkan dalam gambar diagram alir berikut:




**Gambar 3.1 Diagram Alir Penelitian**

# BAB IV

# HASIL DAN PEMBAHASAN

Pada bab ini akan dilakukan kajian untuk mendapatkan estimasi parameter model *small area estimation* dengan pendekatan regresi semiparametrik *penalized* spline beserta aplikasinya untuk menduga pengeluaran per kapita di Kabupaten Sleman pada tahun 2014

## 4.1 Estimator Small Area Estimation dengan Pendekatan Semiparametrik Penalized Spline

Dimisalkan $\theta$ merupakan vektor dari parameter *small area* estimation yang berukuran $m\times1$ dan diasumsikan vektor tersebut merupakan estimator langsung $\hat{\theta}$. Jika dinyatakan $m\times q$ adalah matriks dari variabel penyerta dari level area $\boldsymbol{x}_i = (\boldsymbol{x}_{1i,}\boldsymbol{x}_{2i,\dots}\boldsymbol{x}_{pi})^{\boldsymbol{T}}$ sehingga model SAE berbasis area dapat ditulis seperti persamaan (2.3) adalah sebagai berikut:

$$\boldsymbol{\theta}_i = \boldsymbol{z}_i^T\boldsymbol{\alpha} + b_i\boldsymbol{v}_i + \boldsymbol{e}_i\ ; \quad i = 1,2,\dots,m\ ;\ v_i \sim N(0,\sigma_v^2) \tag{4.1}$$

Dimana

$\boldsymbol{\theta}_i =$ parameter *small area*, berukuran $1\times1$

$\boldsymbol{z}_i = (\boldsymbol{z}_{1i,}\boldsymbol{z}_{2i,\dots}\boldsymbol{z}_{pi})^{\boldsymbol{T}}$ vektor variabel penyerta

$\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_p)^T$ adalah vektor parameter *fixed*

$b_i$ = konstanta positif yang diketahui

$v_i$ = pengaruh acak spesifik yang diasumsikan memiliki distribusi normal $v_i \sim N(0,\sigma_v^2)$

$m =$ jumlah *small area*

Kemudian pada kasus *small area* dengan menggunakan pendekatan *general linear mixed model*, notasi pada model *small area* dapat ditulis sebagai berikut:

$$\boldsymbol{y}_i = \widehat{\boldsymbol{\theta}}_i,\ \boldsymbol{X}_i = z_i^T,\ \mathbf{v}_i = \boldsymbol{v}_i,\ \boldsymbol{e}_i = e_i,\ \boldsymbol{\alpha} = (\alpha_1,\cdots,\alpha_p)^T$$

Sehingga model *small area* yang terbentuk adalah sebagai berikut:

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\alpha} + \boldsymbol{b}_i\mathrm{v_i} + \boldsymbol{e}_i \tag{4.2}$$

Pada model *small area estimation* berbasis area ini akan menghasilkan estimasi *small area* yang *reliable* dengan meminjam kekuatan dari domain lain/variabel penyerta. Pada penelitian ini model *small area* yang akan dibentuk akan menggunakan pendekatan semiparametrik *penalized* spline. Model semiparametrik dengan satu respon secara umum dapat ditulis

$y_i = m(x_i) + e_i\,, i = 1, 2, \ldots, n$

Dengan $m(x_i)$ adalah fungsi yang smooth dimana tidak diketahui bentuknya. Variabel $x_i$ sebagai variabel prediktor dengan $e_i \sim N(0, \sigma^2)$, sehingga diberikan fungsi spline adalah sebagai berikut:

$$\begin{aligned} m(x_1; \beta) &= \beta_0 + \beta_i x_1 + \cdots + \beta_p x_1^p + \textstyle\sum_{j=1}^{k} \gamma_j (x_i - k_j)_+^p \\ &= \boldsymbol{X\beta} \end{aligned}$$

Dengan $p$ adalah derajat spline, $(x_i - k_j)_+ = maks\,\{0, (x_i - k_j)\}, k_j$ dimana $j = 1, \ldots, K$ merupakan himpunan titik knot. Dengan menetapkan $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_p)^T$ adalah $(p+1)$ vektor koefisien fungsi polinomial, $\boldsymbol{\gamma}$=$(\gamma_1, \ldots, \gamma_k)^T$ adalah vektor koefisien spline,

Dengan $(x_i - k_j)_+^p \begin{cases} = (x_i - k_j)_+^p & \text{untuk } x_i \geq k_j \\ = 0 & \text{untuk } x_i < k_j \end{cases}$

Sehingga model dapat ditulis sebagai berikut:

$$y_i = \beta_0 + \beta_i x_i + \cdots + \beta_p x_i^p + \textstyle\sum_{j=1}^{k} \gamma_j (x_i - k_j)_+^p + e_i$$

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Z\gamma} + \boldsymbol{e} \tag{4.3}$$

Menurut Opsomer et al (2008) model spline yang terbentuk akan terindikasi *overparameterized* sehingga akan menyebabkan *overfitting*. Oleh karena itu untuk menghindari hal tersebut ditambahkan penalty pada parameter spline.

Pemilihan kendala ditentukan berdasarkan beberapa kriteria penalty, salah satunya dengan memilih $a$, konstanta nonnegative, sedemikian sehingga $\sum_{i=1}^{K} u_i^2 \leq a$ atau dalam bentuk matriks ditulis menjadi $||\boldsymbol{u}||^2 \leq a$. Kriteria ini dipilih karena mampu mengurangi *overfitting* data (Krivoboka, 2006; Griggs, 2013). Dengan menggunakan metode pengali Langrange, minimisasi problem, estimator *penalized* *spline* diperoleh dengan meminimumkan *penalized least square* $\min_{\beta,\gamma} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\boldsymbol{\gamma}\|^2$ terhadap kendala $\|\gamma\|^2 \leq a$ dapat ditulis menjadi

$$\min_{\beta,\gamma,\lambda}\{\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\boldsymbol{\gamma}\|^2 + \lambda\boldsymbol{\gamma}^T\boldsymbol{\gamma}\} = \min_{\theta,\lambda}\{\|\mathbf{y} - \mathbf{C}\ddot{\theta}\|^2 + \lambda\ddot{\theta}^T\boldsymbol{D}\ddot{\theta}\text{ , } \lambda \geq 0 \qquad (4.4)$$

Dimana memisalkan $\boldsymbol{C} = [\boldsymbol{X}, \boldsymbol{Z}]$ dan $\ddot{\theta} = \begin{bmatrix}\boldsymbol{\beta}\\ \boldsymbol{\gamma}\end{bmatrix}$ dan nilai diketahui $\mathbf{D}$ merupakan matrik penalty

$$D = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{0}_{(P+1)\times 2} & \mathbf{0}_{(P+1)\times K} \\ \mathbf{0}_{K\times(P+1)} & \boldsymbol{I}_{K\times K} \end{bmatrix}$$

Sehingga didapatkan nilai

$$\|\mathbf{Y} - \mathbf{C}\ddot{\theta}\|^2 = (\mathbf{y} - \mathbf{C}\ddot{\theta})^T(\mathbf{y} - \mathbf{C}\ddot{\theta})$$

$$= (\boldsymbol{y}^T - \ddot{\theta}^T\boldsymbol{C}^T)(\boldsymbol{y} - \mathbf{C}\ddot{\theta})$$

$$= \boldsymbol{y}^T\boldsymbol{y} - 2\boldsymbol{y}^T\mathbf{C}\ddot{\theta} + \ddot{\theta}^T\boldsymbol{C}^T\boldsymbol{C}\ddot{\theta}$$

Dengan fungsi kendala $\ddot{\theta}^T\boldsymbol{D}\ddot{\theta} \leq a$, persamaan langrange yang terbentuk adalah

$$L = \boldsymbol{y}^T\boldsymbol{y} - 2\boldsymbol{y}^T\mathbf{C}\ddot{\theta} + \ddot{\theta}^T\boldsymbol{C}^T\boldsymbol{C}\ddot{\theta} + \lambda\ddot{\theta}^T\boldsymbol{D}\ddot{\theta} \qquad (4.5)$$

Kemudian dengan menggunakan metode *penalized least square* untuk meminimumkan persamaan L adalah sebagai berikut:

Syarat perlu

$$\frac{\partial L}{\partial\ddot{\theta}^T} = 0$$

Sehingga

$$\frac{\partial}{\partial\ddot{\theta}^T}(\boldsymbol{y}^T\boldsymbol{y} - 2\boldsymbol{y}^T\mathbf{C}\ddot{\theta} + \ddot{\theta}^T\boldsymbol{C}^T\boldsymbol{C}\ddot{\theta} + \lambda\ddot{\theta}^T\boldsymbol{D}\ddot{\theta}) = 0$$

$$\frac{\partial}{\partial\ddot{\theta}^T}(\boldsymbol{y}^T\boldsymbol{y}) - \frac{\partial}{\partial\ddot{\theta}^T}(2\boldsymbol{y}^T\mathbf{C}\ddot{\theta}) + \frac{\partial}{\partial\ddot{\theta}^T}(\ddot{\theta}^T\boldsymbol{C}^T\boldsymbol{C}\ddot{\theta}) + \frac{\partial}{\partial\boldsymbol{\theta}^T}(\lambda\ddot{\theta}^T\boldsymbol{D}\ddot{\theta}) = 0$$

$$-2\boldsymbol{C}^T\boldsymbol{y} + 2\boldsymbol{C}^T\boldsymbol{C}\ddot{\theta} + 2\lambda\,\boldsymbol{D}\ddot{\theta} = \mathbf{0}$$

$$(\boldsymbol{C}^T\boldsymbol{C} + \lambda\,\boldsymbol{D})\ddot{\theta} = \boldsymbol{C}^T\boldsymbol{y}$$

$\hat{\hat{\theta}} = (\boldsymbol{C}^T\boldsymbol{C} + \lambda\,\boldsymbol{D})^{-\mathbf{1}}\boldsymbol{C}^T\boldsymbol{y}$

Dengan demikian

$\hat{\boldsymbol{y}} = \boldsymbol{C}\hat{\hat{\theta}}$

$$\hat{\boldsymbol{y}} = \boldsymbol{C}(\boldsymbol{C}^T\boldsymbol{C} + \lambda\,\boldsymbol{D})^{-\mathbf{1}}\boldsymbol{C}^T\boldsymbol{y} \tag{4.6}$$

Berdasarkan uraian di aras, nilai $\hat{\theta}$ bergantung pada parameter smoothing $\lambda$. Jika nilai $\lambda$ besar akan menghasilkan bentuk kurva regresi yang sangat halus. Sebaliknya, jika nilai $\lambda$ kecil akan memberikan bentuk kurva regresi yang sangat kasar.

Pada *small area estimation* ini *penalized* spline akan menjadi efek random pada sehingga menghasilkan

$y_i = m(\boldsymbol{X}_i; \boldsymbol{\beta}) + \boldsymbol{\varepsilon}_i$

$= \boldsymbol{X}_i * \boldsymbol{\beta} + \boldsymbol{\varepsilon}_i$

$$= \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{\gamma} + \boldsymbol{\varepsilon}_i \tag{4.7}$$

Model *penalized spline* (4.7) merupakan efek random pada *small area estimation* berbasis area dengan tujuan bisa mendapatkan estimasi pendugaan area kecil secara semiparametrik berdasarkan *linear mixed model.* Dari persamaan (4.2) dan persamaan (4.6) didapatkan model semiparametrik Fay-Herriot dapat ditulis sebagai berikut:

$$\hat{\theta} = \begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{X_1} \end{bmatrix} [\alpha, \beta] + \boldsymbol{Z}\gamma + \boldsymbol{b}v + e$$

Persamaan diatas dapat juga ditulis sebagai berikut:

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{Z\gamma} + \boldsymbol{bv} + \boldsymbol{e} \tag{4.8}$$

Dimana model tersebut terdiri dari fungsi spline yang merupakan semiparametrik $\boldsymbol{X\beta} + \boldsymbol{Z\gamma}$ dan pengaruh acak area kecil $\boldsymbol{bv}$**.** Bentuk matriknya adalah sebagai berikut:

$m$ adalah small area, $v_{1,} \dots, v_T$ merupakan parameter yang diestimasi, kemudian $b_{it} = I\{i\epsilon v_t\}, b_i = (b_i, \dots, b_{ir})$

$$Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}; X = \begin{bmatrix} 1 & x_1 & \dots & x_1^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \dots & x_n^p \end{bmatrix}; , \boldsymbol{Z} = \begin{bmatrix} (x_1 - k_1)_+^p & \dots & (x_1 - k_K)_+^p \\ \vdots & \ddots & \vdots \\ (x_n - k_1)_+^p & \cdots & (x_1 - k_K)_+^p \end{bmatrix}$$

Dimana $(x_1 - k_1)_+^p \begin{cases} (x_1 - k_1)_+^p & \text{untuk } x_i \geq k_j \\ = 0 & \text{untuk } x_i < k_j \end{cases}$

$\gamma \sim \left(0, \Sigma_\gamma\right)$ dengan $\Sigma_\gamma = \sigma_\gamma^2$

$v \sim (0, \Sigma_v)$ dengan $\Sigma_v = \sigma_v^2$

$e \sim (0, \Sigma_\varepsilon)$ dengan $\Sigma_e = \sigma_e^2$ (4.9)

Estimasi pengaruh tetap $\beta$ dapat dilakukan dengan metode *maximum likelihood* (ML) dengan menggangap $\boldsymbol{\gamma}$ dan $\boldsymbol{v}$ sebagai pengaruh acak. Model diatas dapat ditulis menjadi

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{e}^*$$

$$e^* = \boldsymbol{Z\gamma} + \boldsymbol{bv} + \boldsymbol{e}$$

$$\begin{aligned} Var(\boldsymbol{Y}) &= Var(\boldsymbol{X\beta} + \boldsymbol{e}^*) \\ &= Var(\boldsymbol{e}^*) \\ &= Var(\boldsymbol{Z\gamma} + \boldsymbol{bv} + \boldsymbol{e}\ ) \\ &= \boldsymbol{Z}\ \boldsymbol{Var}\ (\boldsymbol{\gamma})\boldsymbol{Z^T} + \boldsymbol{b}\ \boldsymbol{Var}\ (\boldsymbol{v})\boldsymbol{b^T} + \boldsymbol{Var}\ (\boldsymbol{e}) \\ &= \boldsymbol{Z}\ \Sigma_\gamma \boldsymbol{Z^T} + \boldsymbol{b}\Sigma_v \boldsymbol{b^T} + \Sigma_e \\ &= \boldsymbol{V} \end{aligned}$$

atau diperoleh

$$Var(\boldsymbol{Y}) = \boldsymbol{Z}\ \Sigma_\gamma \boldsymbol{Z^T} + \boldsymbol{b}\Sigma_v \boldsymbol{b^T} + \Sigma_e \tag{4.10}$$

Dimana **V** merupakan matriks varians covarians dari $Y$.

Dengan menggangap nilai $\boldsymbol{V}$ diketahui maka fungsi *likelihood* (L) pada persamaan (4.10) adalah:

$$L(\boldsymbol{\beta}|\boldsymbol{Y}) = 2\pi^{\frac{-\sum_{h=1}^{m} n_h}{2}}(|\boldsymbol{V}|)^{1/2}\left[\exp(-\frac{1}{2}(\boldsymbol{Y}-\boldsymbol{X\beta})^T\boldsymbol{V}^{-1}(\boldsymbol{Y}-\boldsymbol{X\beta})\right]$$

Log likelihoodnya adalah

$$\ln L(\boldsymbol{\beta}|\boldsymbol{Y}) = \frac{-\sum_{h=1}^{m} n_h}{2}\ln(2\pi) - \frac{1}{2}\ln(|\boldsymbol{V}|)\text{-}\frac{1}{2}(\boldsymbol{Y}-\boldsymbol{X\beta})^T\boldsymbol{V}^{-1}(\boldsymbol{Y}-\boldsymbol{X\beta})$$

Untuk memperoleh estimasi parameter $\beta$ adalah dengan cara memaksimumkan fungsi likelihood dengan cara menurunkannya terhadap $\beta$

$$\frac{\partial \ln L(\boldsymbol{\beta}|\boldsymbol{Y})}{\partial \boldsymbol{\beta}^T} = -\frac{1}{2}\frac{\partial(\boldsymbol{Y}-\boldsymbol{X\beta})^T\boldsymbol{V}^{-1}(\boldsymbol{Y}-\boldsymbol{X\beta})}{\partial \boldsymbol{\beta}^T}$$

$$=-\frac{1}{2}\frac{\partial(\boldsymbol{Y}^T\boldsymbol{V}^{-1}\boldsymbol{Y}-2\boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{Y}+\boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X\beta})}{\partial \boldsymbol{\beta}^T}$$

$$=\text{-}\frac{1}{2}(-2\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{Y} + 2\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X\beta}) = 0$$

Maka $\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X\beta} = \boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{Y}$

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{Y} \qquad (4.11)$$

Estimasi efek random karena sifatnya yang random dan bukan deterministik, walaupun dalam beberapa literatur perbedaan ini tidak dipermasalahkan prediktor terbaik dari $\boldsymbol{u}$ adalah prediktor yang meminimumkan MSE-nya. Prediktor terbaik dari suatu vektor acak $\boldsymbol{\gamma}$. Jika $\widehat{\boldsymbol{\gamma}}$ adalah prediktor terbaik dari $\gamma$ maka MSE dari prediktor ini adalah:

$$E[(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})^T\boldsymbol{A}(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})] = \iint[(\boldsymbol{\gamma}-\boldsymbol{\gamma})^T A(\boldsymbol{\gamma}-\boldsymbol{\gamma})f(\boldsymbol{\gamma},\boldsymbol{y})]d\boldsymbol{y}d\boldsymbol{\gamma} \qquad (4.12)$$

Dengan $\boldsymbol{A}$ matriks simetris definit positif dengan $f(\boldsymbol{\gamma},\boldsymbol{y})$ adalah fungsi kepadatan peluang gabungan $\boldsymbol{\gamma}$ dan $\boldsymbol{y}$. Dengan meminimumkan persamaan diatas maka diperoleh prediktor terbaik untuk $\gamma$ yang merupakan prediktor tak bias pada sampel $y$ adalah

$\widehat{\boldsymbol{\gamma}} = E(\boldsymbol{\gamma}|\boldsymbol{y})$

$= \boldsymbol{\mu}_{\gamma} + cov\ (\boldsymbol{\gamma}, \boldsymbol{y}^T)\big(cov(\boldsymbol{y})\big)^{-1}(\boldsymbol{y} - \boldsymbol{\mu}_y)$

$= 0 + \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}\ \boldsymbol{Z}^T\boldsymbol{V}^{-1}\big(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\big)$

$$\widehat{\boldsymbol{\gamma}} = \Sigma_{\gamma}\ \boldsymbol{Z}^T\boldsymbol{V}^{-1}\big(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\big) \quad (4.13)$$

Prediktor diatas merupakan prediktor linear tak bias terbaik (BLUP) dari $\boldsymbol{\gamma}$, dimana linear dalam artian linear dalam $y$, tak bias artinya $E(\widehat{\boldsymbol{\gamma}}) = E(\boldsymbol{\gamma})$, dan terbaik artinya prediktor yang meminimumkan MSEnya.

Hal yang sama dilakukan untuk mendapatkan prediktor untuk $\hat{u}$

$$\widehat{\boldsymbol{v}} = \boldsymbol{\Sigma}_{\boldsymbol{v}}\boldsymbol{Z}^T\boldsymbol{V}^{-1}\big(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\big) \quad (4.14)$$

Anggap akan mengestimasi

$$\overline{\boldsymbol{y}}_t = \overline{\boldsymbol{X}}_t\boldsymbol{\beta} + \overline{\boldsymbol{Z}}_t\boldsymbol{\gamma} + \boldsymbol{V}_t \quad (4.15)$$

Dimana $\bar{X}_t$ adalah nilai rata-rata dari $X_i$, $\bar{Z}_t$ adalah basis fungsi spline, dan $V_t$ adalah efek acak small area dengan $V_t = \overline{\boldsymbol{d}}_t\boldsymbol{v} = \boldsymbol{e}_t\boldsymbol{v}$, dan $\boldsymbol{e}_t$ adalah vektor dengan nilai 1 saat ke-t, dan bernilai 0 untuk t lainnya, sehingga prediktor untuk $\hat{y}_t$ yaitu:

$$\widehat{\boldsymbol{y}}_t = \overline{\boldsymbol{X}}_{\boldsymbol{t}}\widehat{\boldsymbol{\beta}} + \overline{\boldsymbol{Z}}_{\boldsymbol{t}}\widehat{\boldsymbol{\gamma}} + \boldsymbol{e}_t\widehat{\boldsymbol{v}}_t \quad (4.16)$$

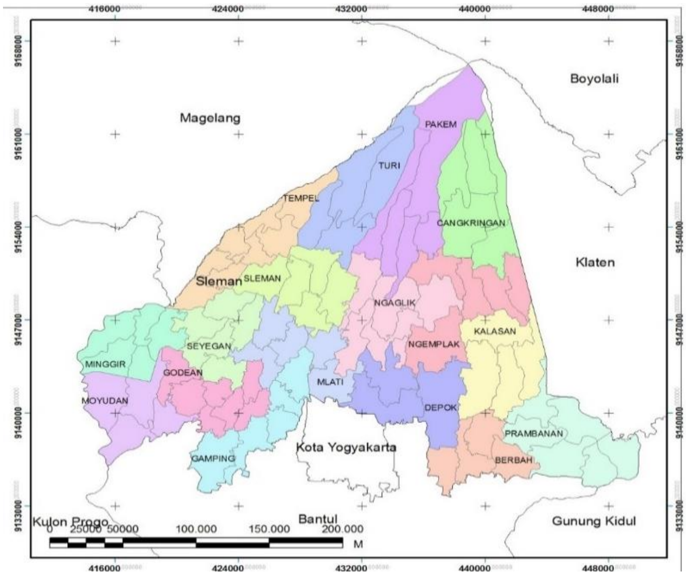
Jika komponen varians tidak diketahui, maka estimator $\beta$ dan prediktor $\gamma$ diperoleh, selanjutnya adalah mengestimasi komponen varians V= $\boldsymbol{Z}\ \Sigma_{\gamma}\boldsymbol{Z}^{\boldsymbol{T}} + \boldsymbol{b}\Sigma_v\boldsymbol{b}^{\boldsymbol{T}} + \Sigma_e$. Estimasi komponen varians adalah $\sigma_{\gamma}^2$, $\sigma_v^2$, dan $\sigma_e^2$. Karena estimasi komponen varians berdasarkan ML bias, maka digunakan metode REML.

Estimasi komponen varianss dengan REML (*Restricted Maksimum Likelihood*) didasarkan pada residual yang dihitung setelah $\beta$ dihitung. Untuk mengestimasi komponen varians dengan REML dapat digunakan beberapa metode, yaitu metode *Newton Rhapson* dan algoritma EM (*expectation and maximization*).

## 4.2 Aplikasi Pemodelan Pengeluaran Per Kapita dengan *Small Area Estimation* dengan Semiparametrik Penalized Spline

### 4.2.1. Deskripsi Umum Kabupaten Sleman

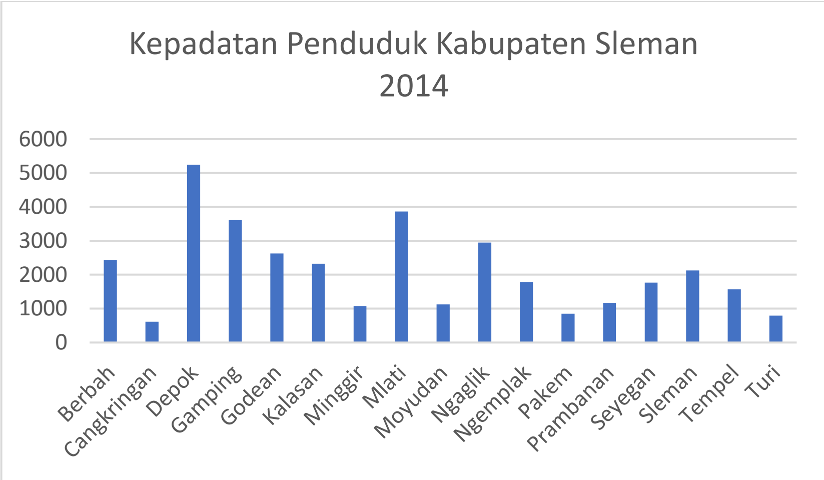
Kabupaten Sleman merupakan salah satu Kabupaten yang terdapat di D.I. Yogyakarta. Secara Geografis Kabupaten Sleman terletak diantara 110° 33′ 00″ dan 110° 13′ 00″ Bujur Timur, 7° 34′ 51″ dan 7° 47′ 30″ Lintang Selatan. Wilayah Kabupaten Sleman sebelah utara berbatasan dengan Kabupaten Boyolali, Propinsi Jawa Tengah, sebelah timur berbatasan dengan Kabupaten Klaten, Propinsi Jawa Tengah, sebelah barat berbatasan dengan Kabupaten Kulon Progo, Propinsi DIY dan Kabupaten Magelang, Propinsi Jawa Tengah dan sebelah selatan berbatasan dengan Kota Yogyakarta, Kabupaten Bantul dan Kabupaten Gunung Kidul, Propinsi D.I.Yogyakarta, seperti pada gambar 4.1




Gambar 4.1 Kabupaten Sleman

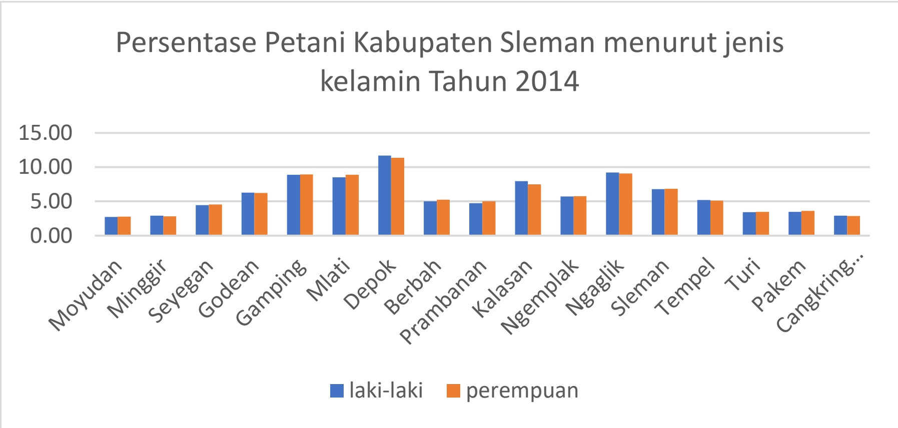
### 4.2.2. Eksplorasi Data Kabupaten Sleman

Menurut Badan Pusat Stastistik (BPS) kepadatan penduduk kabupaten Sleman jumlah penduduk Sleman tahun 2014 sebesar 1.163.970 jiwa, terdiri dari 583.195 laki-laki dan 580.775 perempuan. Dengan luas wilayah 574,82 km2, maka kepadatan penduduk Kabupaten Sleman adalah 2.025 jiwa per $km^2$. Grafik kepadatan penduduk Kabupaten Sleman tahun 2014 dapat dilihat pada gambar 4.2




Gambar 4.2 Kepadatan Penduduk Kabupaten Sleman 2014

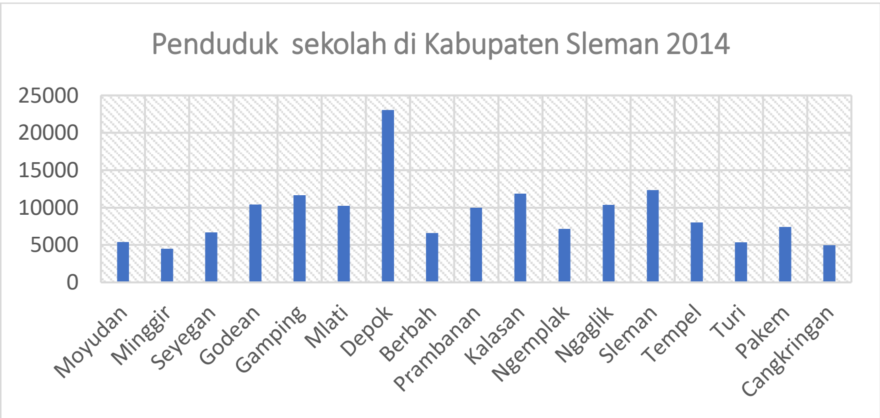
Pada gambar 4.2 mengambarkan ada beberapa kecamatan yang relatif padat penduduknya adalah Depok dengan 5.244 jiwa per $km^2$, Mlati dengan 3.867 jiwa per $km^2$ serta Gamping dan Ngaglik dengan masing-masing 3.608 jiwa dan 2.950 jiwa per $km^2$, kepadatan penduduk di Kabupaten Sleman ini berpengaruh terhadap jumlah anggota keluarga di suatu rumah tangga, menurut harian republika jumlah penduduk Kabupaten Sleman 2013 bertambah dengan jumlah jiwa dalam satu keluarga rata-rata 3,7 jiwa. Jumlah tersebut meningkat dari tahun sebelumnya sebanyak 3,6 jiwa. Oleh karena itu, upaya pemerintah Kabupaten Sleman pada tahun 2014 untuk menggalakan program Keluarga Berencana (KB) untuk menekan jumlah penduduk sudah semakin banyak, berita yang dikutip pada Slemankab (2013) menjelaskan bahwa laju pertambahan penduduk tersebut merupakan angka yang cukup tinggi sehingga merupakan tanggungjawab pemerintah untuk mengendalikan jumlah penduduk terutama angka kelahiran. Salah satu upaya untuk mengendalikan jumlah penduduk tersebut adalah dengan meningkatkan dan memantapkan pelaksanaan keluarga berencana. Jumlah penduduk yang besar ini disatu sisi dapat merupakan potensi SDM yang kaya, namun bila tidak diimbangi dengan peningkatan kesejahteraan masyarakat justru akan menambah beban kerja pemerintah daerah dan apabila pertambahan penduduk tidak terkendali maka masalah yang akan timbul lebih besar, oleh karena itu program KB harus dilaksanakan dengan baik, tahun 2014 jumlah petani yang di Kabupaten Sleman berjenis kelamin laki-laki berjumlah 48.331 orang dan yang berjenis kelamin perempuan berjumlah 45.484. Persentase jumlah petani tiap kecamatan di Kabupaten Sleman pada tahun 2014 terdapat pada gambar 4.3




Gambar 4.3 Persentase Petani Kabupaten Sleman 2014

Persentase petani yang tertinggi terdapat pada beberapa kecamatan diantara lain kecamatan Depok, dengan persentase untuk laki-laki adalah sebesar 11.68% dan perempuan adalah sebesar 11.37%, kemudian kecamatan Ngaglik adalah sebesar 9.23% untuk laki-laki dan 9.06 % untuk perempuan, dan kecamatan Gamping persentase yang berjenis kelamin laki-laki merupakan petani adalah sebesar 8.91 dan untuk yang berjenis kelamin perempuan adalah sebesar 8.92 persen.

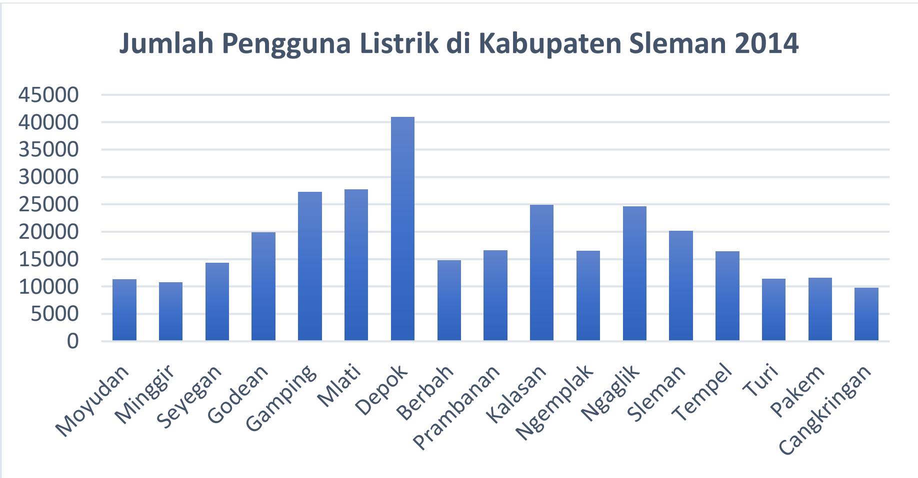
Kemudian eksplorasi data dilanjutkan pada jumlah penduduk yang bersekolah, pada penelitian ini jumlah penduduk yang bersekolah di Kabupaten Sleman di batasi dari SD hingga SMA/SMK. Jumlah penduduk yang bersekolah dapat dilihat pada gambar 4.4




Gambar 4.4 Penduduk sekolah di Kabupaten Sleman 2014

Pada gambar 4.4 dapat dijelaskan bahwa jumlah kecamatan yang tertinggi terdapat pada kecamatan Depok yang berjumlah 23.055 dengan jumlah murid SD adalah sebesar 12.618, jumlah murid  SMP adalah sebesar 3.759, jumlah murid SMA adalah sebesar 1914 orang, dan jumlah murid SMK adalah sebesar 4.764, kemudian kecamatan Sleman berjumlah 12.350 orang dengan jumlah murid SD adalah sebesar 6884 orang, jumlah murid SMP adalah sebesar 2.955 orang, jumlah murid SMA adalah sebesar 1.043, orang dan jumlah murid SMK adalah sebesar 1.468 orang.

Penyediaan tenaga listrik bertujuan untuk meningkatkan perekonomian serta memajukan kesejahteraan masyarakat. Bila tenaga listrik telah dicapai pada suatu daerah atau wilayah maka kegiatan ekonomi dan kesejahteraan pada daerah tersebut dapat meningkat.  Pada tahun 2014 jumlah pengguna listrik di Kabupaten Sleman dapat dijelaskan pada gambar 4.5.



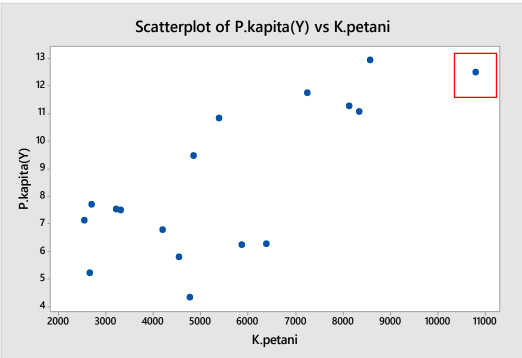
Gambar 4.5 Jumlah Pengguna Listrik di Kabupaten Sleman Tahun 2014

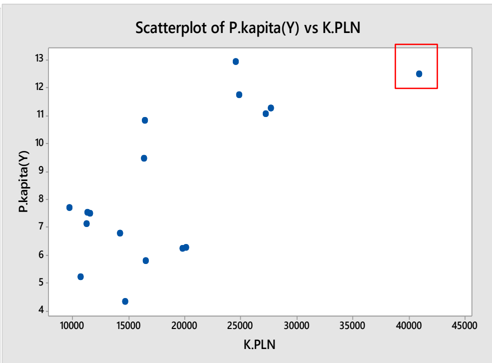
Jumlah pengguna listrik di Kabupaten Sleman tahun 2014 tertinggi terdapat pada kecamatan Depok 40.956, kemudian kecamatan Mlati adalah sebesar 27.717 dan kecamatan Gamping adalah sebesar 27.266 pengguna listrik.

### 4.2.3 Eksplorasi Data Hubungan Antar Variabel

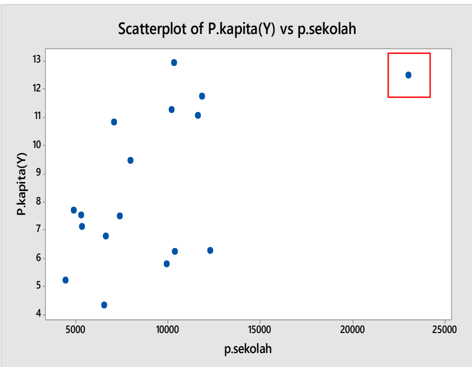
Pada sub bab sebelumnya sudah telah dijelaskan secara deskripsi tiap variabel yang digunakan dalam penelitian ini. Untuk melakukan pemodelan *small area estimation* dengan pendekatan semiparametrik *penalized* spline akan dimulai dengan melakukan eksplorasi pola hubungan antara variabel respon dengan variabel penyerta yang disajikan pada *scatterplot*, eksplorasi ini juga menentukkan variabel parametrik dan variabel nonparametrik.

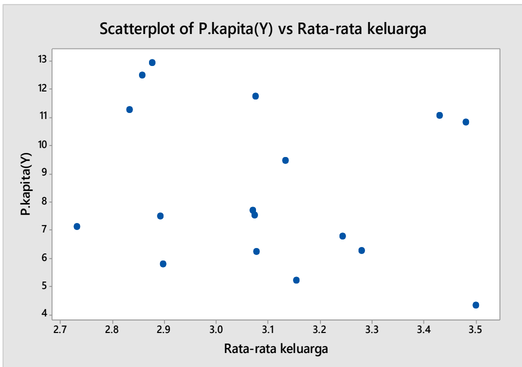
Gambar 4.7 merupakan hasil *scatterplot* antara pengeluaran per kapita dengan jumlah keluarga petani (a), pengeluaran per kapita dengan keluarga pengguna listrik PLN (b), pengeluaran per kapita dan jumlah penduduk yang sedang sekolah (c), serta pengeluaran per kapita dan rata-rata angggota keluarga (d) di Kabupaten Sleman. Gambar 4.6 (a), (b), dan (c) menunjukkan adanya pola hubungan antara liniear yang mengikuti garis lurus, dengan kelinearan yang menunjukkan arah kanan, dimana artinya hubungan kelinearan tersebut bersifat positif. Pada Gambar 4.6 (d) menunjukkan adanya pola hubungan yang menunjukkan kelinearan yang mengarah ke kiri dimana hasil kelinearannya bersifat negative, keempat variabel tersebut merupakan variabel yang bersifat parametrik, hal ini beralasan karena keempat variabel tersebut terbentuk pola data yang sama.




(a) Jumlah Keluarga Petani (b) Keluarga Pengguna Listrik PLN

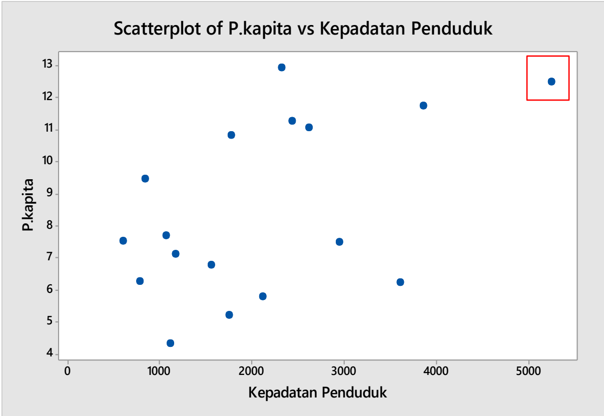



(c ) Jumlah Penduduk yang Sedang Sekolah (d) Rata-Rata Anggota Keluarga

Gambar 4.6 Scatterplot Empat Variabel

Kemudian akan disajikan *scatterplot* antara pengeluaran per kapita dengan jumlah penduduk di Kabupaten Sleman tahun 2014 pada gambar 4.7




Gambar 4.7 Scatterplot Kepadatan Penduduk

Gambar 4.7 merupakan *scatterplot* antara pengeluaran per kapita dengan kepadatan penduduk ($x_5$) pola hubungan yang terbentuk tidak mengikuti pola kelinearan atau mempunyai pola data tertentu dimana perilaku pola data ini dapat berubah-ubah, sehingga variabel pada gambar 4.8 merupakan variabel nonparametrik. Gambar 4.6 (a), (b), (c), dan gambar 4.7 menunjukkan adanya kecamatan yang memiliki nilai yang ekstrem, kecamatan tersebut yaitu kecamatan Depok. Kecamatan Depok ini menurut berita yang dikutip dari Tribun Jogja (2015) merupakan wilayah dengan jumlah penduduk terpadat dimana pertumbuhan ekonomi berkembang pesat, kecamatan Depok ini berbatasan langsung dengan pusat kota sehingga menjadi penyangga Kota Yogyakarta yaitu sebagai daerah tujuan untuk melanjutkan pendidikan, dan daerah pengembangan pemukiman/perumahan, kemudian dari sisi penggunaan listrik PLN adanya peningkatan pemukiman hal tersebut berbanding dengan pengeluaran per kapita yang di keluarkan dan menyebabkan konsumsi listrik di Kecamatan Depok mengalami peningkatan dan menjadi konsumsi tertinggi pada sektor rumah tangga. Semakin tingginya pendapatan menyebabkan pengeluaran per kapita yang dikeluarkan menjadi meningkat.

### 4.2.4 Pemodelan Pengeluaran Per Kapita dengan Menggunakan Semiparametrik Penalized Spline

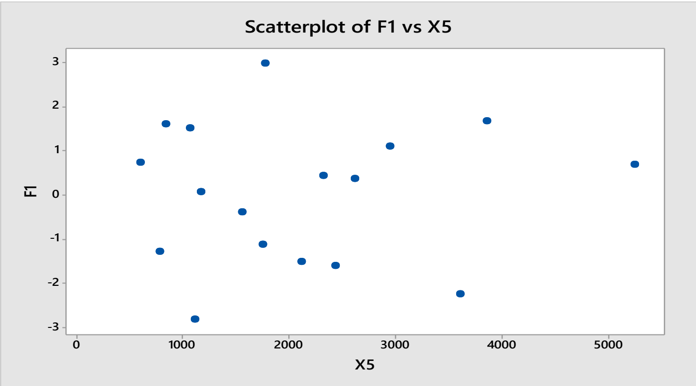
Pada penelitian ini akan dibagi menjadi dua tahapan dalam pembentukkan model dengan menggunakan metode *Generalized Additive Model*, bentuk model untuk semiparametrik *penalized* spline dengan menggunakan pendekatan *Generalized Addittive Model* dapat ditulis sebagai berikut:

$$y = f(x_1, x_2, x_3, x_4) + f(x_5) + v \quad (4.17)$$

Proses pembentukkan model dengan *generalized additive model* ini akan dibagi menjadi dua tahapan, yaitu mencari nilai parametrik untuk $f(parametrik)$ sehingga menghasilkan nilai parametrik adalah sebagai berikut:

$$y = 12.38 + 0.000928x_1 + 0.000197x_2 - 0.000476x_3 - 2.70x_4 + f(x_5) + v \quad (4.18)$$

Nilai residual untuk yang bernilai parametrik dapat dilihat pada lampiran 3, kemudian akan digunakan untuk tahapan pencarian nilai nonparametrik *penalized* spline untuk $f(x_5)$. Pola hubungan antara residual parametrik dengan variabel nonparametrik disajikan pada gambar 4.9




Gambar 4.8 Scatterplot antara Residual $f_1$ dan $x_5$

Gambar 4.8 merupakan *scatterplot* yang dihasilkan antara residual nilai parametrik dengan variabel kepadatan penduduk $(x_5)$ terlihat bahwa pola data yang terbentuk tidak membentuk hubungan tertentu, sehingga gambar 4.8 bersifat nonparametrik, dengan yang tidak mempunyai pola tertentu tentu saja dapat didekati dengan menggunakan spline mempunyai model polynomial tersegmen yang dibangun sedemikian rupa sehingga kurva yang dibentuk mulus pada titik-titik yang disebut knot, dimana dua polinomial disatukan, akan tetapi spline mempunyai kekurangan yaitu jika knot yang dihasilkan terlalu banyak akan mengakibatkan model tersebut menjadi *overfitting*. Untuk mengatasi hal tersebut penelitian ini menggunakan pendekatan *penalized* spline mendapatkan nilai nonparametrik, kelebihan pada p*enalized* spline ini menurut Hall dan Opsomer (2005) dapat mengatasi model yang *overfitting* jika jumlah knot yang digunakan terlampau banyak dengan menambahkan penalty/kendala pada parameter spline dengan tujuan untuk menghindari kelebihan knot, proses tersebut terdapat pada persamaan (4.4). Proses nonparametrik *penalized spline*, terlebih dahulu dilakukan pemilihan jumlah dan lokasi knot. Pemilihan jumlah knot optimum digunakan *fixed selection method*, yaitu

$$\mathrm{K} = \min(\tfrac{1}{4} \times \text{banyaknya } x_i \text{yang } unique, 35)$$

Karena nilai $x_i$ yang digunakan *unique*, artinya setiap data $x_i$ tidak ada yang bernilai sama. Dengan demikian jumlah knot yang dapat dihitung sebagai berikut:

$$\mathrm{K} = \min(\tfrac{1}{4} \times n, 35)$$

$$= \min(\tfrac{1}{4} \times 17{,}35)$$

$$= \min(4.25{,}35)$$

$$= 4.25 \approx 4$$

Artinya, jumlah knot yang akan digunakan dalam penelitian ini adalah 1,2,3, dan 4. Kemudian menghitung nilai GCV untuk masing-masing knot yang telah ditentukkan berdasarkan *fixed selection method*. Nilai GCV minimum yang diperoleh adalah 2.96563. Berdasarkan kriteria GCV minimum diperoleh knot optimum sebanyak 4 buat knot yang masing-masing terletak pada titik 1078.333, 1585.857, 2151.143, 2688.524. Perhitungan nilai λ optimum bertujuan untuk mengetahui kurva yang terbentuk dari regresi *penalized* spline, pada penelitian ini diperoleh nilai λ optimum, sebesar 1.00216e+13. Model semiparametrik *penalized* spline yang terbentuk dengan proses *generalized additive model* adalah sebagai berikut:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \gamma_1 (x_5 - 10788.33)_+ + \gamma_2 (x_5 - 1585.85)_+ + \gamma_3 (x_5 - 2151.14)_+ + \gamma_4 (x_5 - 2688.52)_+ + v \quad (4.19)$$

Hasil pada persamaan (4.19) merupakan hasil knot yang didapatkan dengan menggunakan *penalized* spline, kemudian persamaan (4.19) juga merupakan model yang akan digunakan untuk menduga pengeluaran per kapita di Kabupaten Sleman pada level area. Pada model semiparametrik *penalize*d spline ini, nilai $\beta$ didapatkan dengan memaksimumkan fungsi *likelihood* atau *loglikeliho*od sehingga didapatkan nilai estimasi $\beta$ pada tabel 4.1

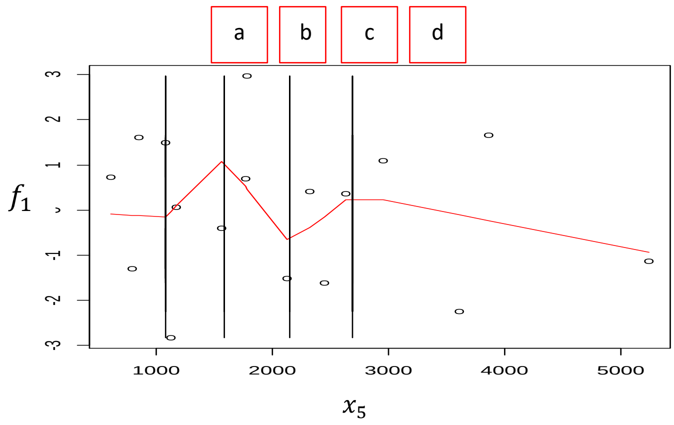
Tabel 4.1 Nilai estimasi $\beta$

| **Parameter** | **Estimasi** |
|---|---|
| $\hat{\beta}_0$ | 12.37 |
| $\hat{\beta}_1$ | 0.000928 |
| $\hat{\beta}_2$ | 0.000197 |
| $\hat{\beta}_3$ | -0.000476 |
| $\hat{\beta}_4$ | -2.69895 |

Tabel 4.1 menunjukkan hasil nilai estimasi, sehingga nilai $\beta$ untuk untuk model semiparametrik yang terbentuk adalah sebagai berikut:

$$y = 12.37 + 0.000928 x_1 + 0.000197 x_2 - 0.000476 x_3 - 2.69895 x_4 + \gamma_1 (x_5 - 1078.33)_+ \gamma_2 (x_5 - 1585.85)_+ + \gamma_3 (x_5 - 2151.14)_+ + \gamma_4 (x_5 - 2688.52)_+ + v \quad (4.20)$$

Interpretasi pada model (4.20) dengan menunjukkan bahwa terdapat perbedaan hasil untuk setiap kecamatan, kemudian titik lokasi knot yang dihasilkan dengan menggunakan *penalized* spline disajikan pada gambar 4.9




Gambar 4.9 Lokasi Titik Knot

a. Titik knot (a) yang terdapat pada gambar 4.9 merupakan nilai knot yang bernilai 1078.333 yang artinya jika nilai $x_5 \leq 1078.33$, setiap kenaikan satu satuan akan berpengaruh sebesar (0.000928$x_1$+0.000197$x_2$-0.000476$x_3$-2.69895$x_4 + \gamma_1$) satuan terhadap $y$
b. Titik knot (b) yang terdapat pada gambar 4.9 menunjukkan nilai knot yang bernilai 1585.857 yang artinya jika nilai $1078.33 < x_5 \leq 1585.85$, setiap kenaikan satu satuan $x_5$ akan berpengaruh sebesar (0.000928$x_1$+0.000197$x_2$-0.000476$x_3$-2.69895$x_4 + \gamma_1 + \gamma_2$) satuan terhadap $y$
c. Titik knot (c) pada gambar 4.9 menunjukkan knot yang bernilai 2151.14 yang artinya jika nilai $1585.85 < x_5 \leq 2151.14$, setiap kenaikan satu satuan $x_5$ akan berpengaruh sebesar (0.000928$x_1$+0.000197$x_2$-0.000476$x_3$-2.69895$x_4 + \gamma_1 + \gamma_2 + \gamma_3$) satuan terhadap $y$
d. Titik knot (d) pada gambar 4.9 menunjukkan knot yang bernilai 2688.52, yang artinya jika nilai 2151.14< $x_5 \leq 2688.52$ sesetiap kenaikan satu satuan $x_5$ akan berpengaruh sebesar (0.000928$x_1$+0.000197$x_2$-0.000476$x_3$-2.69895$x_4 + \gamma_1 + \gamma_2 + \gamma_3 + \gamma_4$) satuan terhadap $y$

Pada persamaan (4.20) jika memasukkan nilai $x_5$ pada setiap knot menghasilkan model yang berbeda-beda untuk setiap kecamatan yang disajikan pada tabel 4.2

Tabel 4.2 Model yang terbentuk untuk setiap Small Area estimation

| No | Kecamatan | Model |
|---|---|---|
| 1 | Berbah | $12.37 + 0.000928x_1$+$0.000197x_2$-$0.000476x_3$-$2.69895x_4$+$46.67\gamma_1 - 460.85\,\gamma_2 - 1026.14\gamma_3 - 1563.52\gamma_4 + v$ |
| 2 | Cangkringan | $12.37 + 0.000928x_1$+$0.000197x_2$-$0.000476x_3$-$2.69895x_4$+$(-2.33)\gamma_1 + (-509.85)\gamma_2 + (-1075.14)\gamma_3 + (-1612)\gamma_4 + v$ |
| 3 | Depok | $12.37 + 0.000928x_1$+$0.000197x_2$-$0.000476x_3$-$2.69895x_4$+$686.67\gamma_1 + 179.15\gamma_2 + (-386.14)\gamma_3 + (-923.52)\gamma_4 + v$ |
| 4 | Gamping | $12.37 + 0.000928x_1$+$0.000197x_2$-$0.000476x_3$-$2.69895x_4$+$1548.67\gamma_1 + 1041.15\gamma_2 + 475.86\gamma_3 + (-61.52)\gamma_4 + v$ |
| 5 | Godean | $12.37 + 0.000928x_1$+$0.000197x_2$-$0.000476x_3$-$2.69895x_4$+$2529.67\gamma_1 + 2022.15\gamma_2 + 1456.86\gamma_3 + 919.48\gamma_4 + v$ |
| 6 | Kalasan | $12.37 + 0.000928x_1$+$0.000197x_2$-$0.000476x_3$-$2.69895x_4$+$2788.67\gamma_1 + 2281.15\gamma_2 + 1715.86\gamma_3 + 1178.48\gamma_4 + v$ |
| 7 | Minggir | $12.37 + 0.000928x_1$+$0.000197x_2$-$0.000476x_3$-$2.69895x_4$+$4165.67\gamma_1 + 3658.15\gamma_2 + 3092.86\gamma_3 + 2555.48\gamma_4 + v$ |
| 8 | Mlati | $12.37 + 0.000928x_1$+$0.000197x_2$-$0.000476x_3$-$2.69895x_4$+$1364.67\gamma_1 + 857.15\gamma_2 + 291.86\gamma_3 + (-245.52)\gamma_4 + v$ |
| 9 | Moyudan | $12.37 + 0.000928x_1$+$0.000197x_2$-$0.000476x_3$-$2.69895x_4$+$97.67\gamma_1 + (-409.85)\gamma_2 + (-975.14)\gamma_3 + (-1512.52)\gamma_4 + v$ |
| 10 | Ngaglik | $12.37 + 0.000928x_1$+$0.000197x_2$-$0.000476x_3$-$2.69895x_4$+$1247.67\gamma_1 + 740.15\gamma_2 + 174.86\gamma_3 + (-362.52)\gamma_4 + v$ |
| 11 | Ngemplak | $12.37 + 0.000928x_1$+$0.000197x_2$-$0.000476x_3$-$2.69895x_4$+$706.67\gamma_1 + 199.15\gamma_2 + (-366.14)\gamma_3 + (-903.52)\gamma_4 + v$ |
| 12 | Pakem | $12.37 + 0.000928x_1$+$0.000197x_2$-$0.000476x_3$-$2.69895x_4$+$1871.67\gamma_1 + 1364.15\gamma_2 + 798.86\gamma_3 + 261.48\gamma_4 + v$ |
| 13 | Prambanan | $12.37 + 0.000928x_1$+$0.000197x_2$-$0.000476x_3$-$2.69895x_4$+$1043.67\gamma_1 + 536.15\gamma_2 + (-29.14)\gamma_3 + (-566.52)\gamma_4 + v$ |
| 14 | Seyegan | $12.37 + 0.000928x_1$+$0.000197x_2$-$0.000476x_3$-$2.69895x_4$+$488.67\gamma_1 + (-18.85)\gamma_2 + (-584.14)\gamma_3 + (-1121.52)\gamma_4 + v$ |
| 15 | Sleman | $12.37 + 0.000928x_1$+$0.000197x_2$-$0.000476x_3$-$2.69895x_4$+(-281.33)$\gamma_1 + (-788.85)\gamma_2 + (-1354.14)\gamma_3 + (-1891.52)\gamma_4 + v$ |
| 16 | Tempel | $12.37 + 0.000928x_1$+$0.000197x_2$-$0.000476x_3$-$2.69895x_4$+(-228.33)$\gamma_1 + (-735.85)\gamma_2 + (-1301.14)\gamma_3 + (-1838.52)\gamma_4 + v$ |
| 17 | Turi | $12.37 + 0.000928x_1$+$0.000197x_2$-$0.000476x_3$-$2.69895x_4$+(-466.33)$\gamma_1 + (-973.85)\gamma_2 + (1539.14)\gamma_3 + (-2076.52)\gamma_4 + v$ |

Model yang terbentuk untuk setiap kecamatan memiliki efek random, seperti yang dijabarkan pada persamaan (4.8) dimana nilai $\gamma$ dan $v$ merupakan efek random karena pada penelitian ini varians/ komponen ragam untuk faktor acak tidak diketahui sehingga harus ditaksir dengan menggunakan metode *maximum likelihood* (ML) akan tetapi jika estimasi komponen varians dengan menggunakan ML bias, maka dapat menggunakan metode *restricted maximum likelihood* (REML). Nilai $\gamma$ dan $v$ merupakan faktor acak dari model semiparametrik *penalized* spline, nilai $\gamma$ dan $v$ merupakan *empirical best linear unbiased predictors* (EBLUP). Metode EBLUP ini merupakan perluasan dari metode BLUP dimana komponen ragam diketahui, metode EBLUP juga mensubstitusikan komponen ragam yang tidak diketahui ke- dalam penduga BLUP. Untuk nilai nilai $\gamma$ dan $v$ efek random terdapat pada lampiran 4.

Pendugaan pengeluaran per kapita di Kabupaten Sleman pada level area dengan menggunakan model semiparametrik *penalized* spline pada tahun 2014 berdasarkan tabel 4.3 didapatkan rata-rata pengeluaran per kapita untuk kabupaten Sleman adalah sebesar Rp 846.891,13 dengan koefisien ragam adalah sebesar 31.51, hal tersebut menunjukkan bahwa nilai pendugaan pengeluaran per kapita di Kabupaten Sleman tidak terlalu bervariasi. Ringkasan hasil pendugaan pengeluaran per kapita disajikan pada tabel 4.3

Tabel 4.3 Ringkasan Pendugaan Pengeluaran Perkapita di Kabupaten Sleman

| **Statistik** | **Pendugaan Pengeluaran Per kapita** |
|---|---|
| Rata-rata | 8.469 |
| Koefisien Ragam | 31.51 |
| Variansi | 7.120 |
| Minimum | 4.323 |
| Maximum | 13.99 |

Sumber: Hasil Olahan Data

Pengeluaran per kapita tertinggi di Kabupaten Sleman pada tahun 2014 terdapat pada kecamatan Ngaglik yaitu sebesar Rp 1.399.921,13 untuk pengeluaran per kapita terendah tahun 2014 di Kabupaten Sleman terdapat pada kecamatan Berbah yaitu sebesar Rp 432,268.87. Hasil pendugaan pengeluaran per kapita untuk seluruh kecamatan disajikan pada tabel 4.4

Tabel 4.4 Hasil Pendugaan Pengeluaran Per Kapita Kabupaten Sleman Level Kecamatan 2014

| **Kecamatan** | **Pengeluaran Per Kapita($\times$ Rp 100000)** |
|---|---|
| Berbah | 4.322687 |
| Cangkringan | 5.049402 |
| Depok | 10.26999 |
| Gamping | 8.436958 |
| Godean | 7.176440 |
| Kalasan | 9.667969 |
| Minggir | 6.390384 |
| Mlati | 13.964470 |
| Moyudan | 7.769069 |
| Ngaglik | 13.999211 |
| Ngemplak | 9.442311 |
| Pakem | 9.363858 |
| Prambanan | 7.668593 |
| Seyegan | 5.550187 |
| Sleman | 9.194890 |
| Tempel | 8.518848 |
| Turi | 7.186200 |

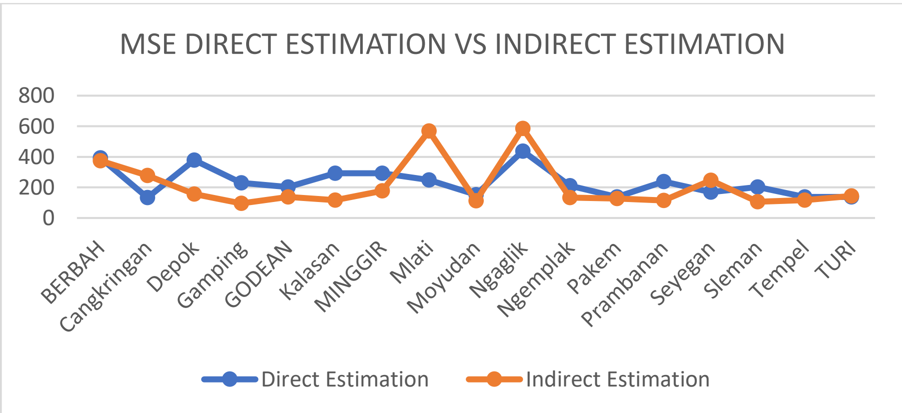
## 4.3 Kebaikan Model Pendugaan

Kebaikan model merupakan salah satu hal yang dilakukan untuk melihat apakah model yang telah terbentuk baik atau tidak, untuk mengetahui model yang telah terbentuk baik atau tidak akan dilakukan pendugaan nilai MSE dengan menggunakan pendugaan tidak langsung (*indirect*) dan langsung (*direct*). Pendugaan tidak langsung pada subbab ini merupakan pendugaan yang menggunakan model yang mempunyai efek *random* untuk mengetahui *small area estimation* yang dihasilkan, sedangkan pada pendugaan langsung untuk menduga suatu area tidak memerlukan model. Kebaikan model pendugaan ini akan menggunakan metode *resampling jackknife* dikarenakan populasi pada parameter untuk data *real* yang belum diketahui sehingga dengan menggunakan *resampling parameter* pada populasi yang akan dicari, untuk proses MSE dengan metode *jackknife* sesuai dengan prosedur yang terdapat pada persamaan (2.21) sehingga diperoleh hasilnya sesuai dengan tabel 4.5

Tabel 4.5 Hasil MSE dengan Jackknife

| Kecamatan | MSE Pendugaan Langsung | MSE Pendugaan Tidak Langsung |
|---|---|---|
| Berbah | 393.81 | 373.45 |
| Cangkringan | 132.96 | 278.17 |
| Depok | 378.70 | 156.04 |
| Gamping | 229.22 | 95.31 |
| Godean | 203.34 | 137.48 |
| Kalasan | 291.45 | 116.34 |
| Minggir | 292.91 | 176.92 |
| Mlati | 247.59 | 568.49 |
| Moyudan | 153.02 | 112.14 |
| Ngaglik | 437.72 | 585.92 |
| Ngemplak | 210.16 | 132.97 |
| Pakem | 138.57 | 126.37 |
| Prambanan | 237.64 | 115.47 |
| Seyegan | 169.29 | 246.33 |
| Sleman | 202.13 | 106.28 |
| Tempel | 137.95 | 116.88 |
| Turi | 137.85 | 142.87 |

Berdasarkan tabel 4.5 terlihat bahwa MSE menggunakan pendugaan langsung dan tidak langsung menghasilkan perbedaan untuk setiap kecamatan. Ada beberapa pendugaan tidak langsung yang menghasilkan nilai MSE yang cukup besar dibandingkan kecamatan lain yaitu pada kecamatan Cangkringan, Kecamatan Mlati, dan Kecamatan Ngaglik. Secara visual hasil perbandingan antar grafik MSE penduga langsung dan penduga tidak langssung disajikan pada gambar 4.10 dan hasil secara deskripsi untuk kedua metode tersebut dapat dilihat pada tabel 4.6

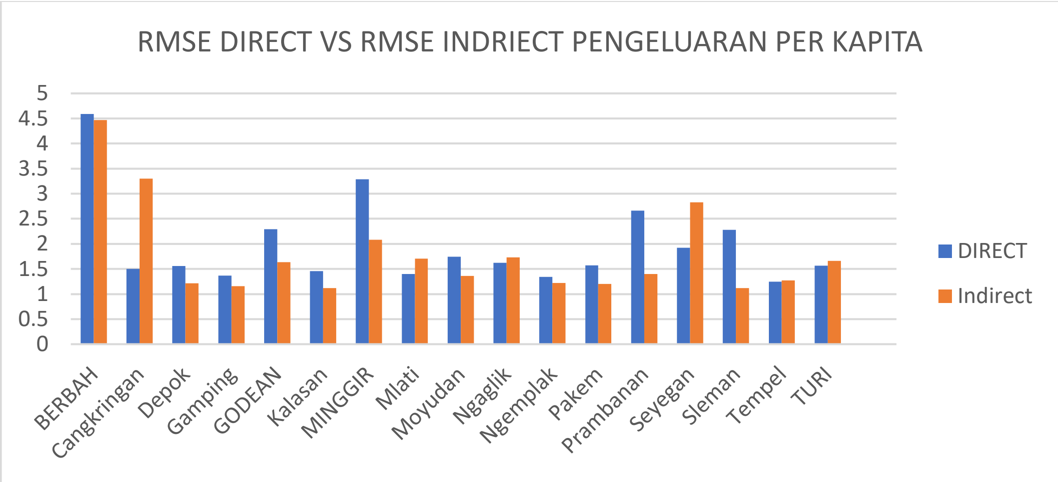



Gambar 4.10 Grafik MSE Pendugaan Langsung dan Tidak Langsung

Tabel 4.6 Ringkasan MSE Jackknife

| Statistik | MSE Pendugaan Langsung | MSE Pendugaan Tidak Langsung |
|---|---|---|
| Rata-rata | 234 | 211.03 |
| Variansi | 8588.78 | 22886 |
| Minimum | 132.96 | 95.31 |
| Maximum | 437.72 | 585.92 |

Berdasarkan Tabel 4.6 nilai rata-rata *jackknife* dengan menggunakan pendugaan langsung didapatkan adalah sebesar 235 dengan variansi adalah sebesar 911 dengan menggunakan pendugaan tidak langsung menghasilkan rata-rata adalah sebesar 211.03 dengan variansi yang didapatkan adalah sebesar 22886, variansi yang dihasilkan oleh pendugaan tidak langsung besar disebabkan adanya beberapa kecamatan yang menghasilkan MSE yang cukup besar hal tersebut juga berpengaruh dari hasil pendugaan pengeluaran per kapita yang dihasilkan pada tabel 4.3 dimana pendugaan pengeluaran per kapita yang dihasilkan untuk Kecamatan Mlati dan Kecamatan Ngaglik cukup besar, akan tetapi secara keseluruhan untuk kebaikan model hasil dengan menggunakan MSE dengan pendugaan tidak langsung mendapatkan nilai yang lebih kecil dibandingkan dengan MSE pendugaan langsung. Gambar 4.10 menggambarkan perbandingan MSE antar kecamatan, terlihat bahwa dengan menggunakan pendugaan tidak langsung menghasilkan grafik yang lebih kecil dibandingkan dengan pendugaan langsung, kemudian seperti yang telah dijelaskan pada tabel 4.5 bahwa terdapat kecamatan yang memiliki hasil MSE tertinggi bukan merupakan daerah yang ekstrem seperti pada gambar 4.7 dan gambar 4.8, dan MSE dengan hasil yang tinggi tersebut disebabkan pendugaan pengeluaran per kapita yang menghasilkan nilai yang besar dibandingkan dengan kecamata lain. Diberikan hasil akar nilai MSE yang disebut dengan RMSE, RMSE yang dihasilkan untuk kedua metode tersebut dapat dilihat pada gambar 4.10




Gambar 4.11 RMSE Pendugaan Langsung dan Tidak Langsung

Gambar 4.11 menunjukkan perbandingan RMSE yang dihasilkan pada metode *small area estimation* dengan mengunakan pendugaan langsung dan metode pendugaan tidak langsung yang didekati dengan *jackknife*, terlihat bahwa RMSE yang dihasilkan dengan menggunakan metode langsung menghasilkan nilai yang lebih kecil, walaupun ada beberapa daerah yang memiliki hasil yang lebih besar. Perbedaan RMSE yang dihasilkan disebabkan nilai untuk pembagi pada pendugaan langsung dan tidak langsung berbeda.

### 4.4 Pembahasan Small Area Estimation dengan Semiparametrik Penalized Spline

Pada subbab ini akan membahas keseluruhan dari materi yang telah di analisis pada *small area estimation* dengan pendekatan semiparametrik *penalized* spline. Metode *Small area estimation* ini merupakan teknik pendugaan tidak langsung di suatu area yang relative kecil yang membutuhkan informasi tambahan yang memiliki hubungan dengan peubah yang sedang diamati. Pada penelitian ini penggunaan *small area estimation* ini menggunakan pendekatan semiparametrik hal ini berdasarkan gambar 4.7 dan gambar 4.8 eksplorasi hubungan antar variabel memiliki variabel yang parametrik dan nonparametrik sehingga penulis menggunakan semiparametrik *penalized* spline. Menurut Giusti et al (2012), penggunaan semiparametrik *penalized* spline berbasis level area menggunakan model Fay-Herriot yang didekati dengan menggunakan *generalized mixed model* sehingga model yang terbentuk seperti pada persamaan (4.2). *Penalized* spline digunakan pada penelitian ini digunakan karena *penalized* spline merupakan salah satu pendekatan *smoothing* yang sederhana dan fleksibel (Opsomer, 2005), kemudian *penalized* spline ini menurut Opsomer (2008) merupakan efek *random* untuk *small area estimation* sehingga model yang terbentuk menurut Giusti et al (2012) sesuai dengan persamaan (4.8). Model ini dapat digunakan pada situasi dimana informasi tersedia hanya pada level area dan hubungan antara variabel penyerta dan respon tidak dapat ditentukkan secara langsung (Giusti, 2012). Model *small area estimation* dengan pendekatan semiparametrik *penalized* spline yang dihasilkan memiliki perbedaan antar kecamatan atau observasi seperti pada tabel 4.2 yang membedakan *small area estimation* dengan model pendugaan langsung yaitu terdapat pada efek random sesuai yang dihasilkan yang terdapat pada lampiran 4 sehingga pengeluaran per kapita yang diduga menjadi berbeda antar kecamatan. Hasil yang didapatkan pada penelitian ini sesuai dengan subab sebelumnya yaitu pendugaan pengeluaran per kapita di Kabupaten Sleman 2014 tertinggi terdapat pada Kecamatan Ngaglik, kecamatan ini merupakan daerah yang langsung berbatasan dengan Kota Yogyakarta, akan tetapi faktor-faktor tertentu dapat menyebabkan pengeluaran per kapita di daerah tersebut menjadi lebih tinggi yaitu adanya pemekaran wilayah sehingga terciptanya lapangan kerja baru yang dapat meningkatkan pendapatan rumah tangga di Kecamatan Ngaglik tersebut. Kebaikan model pada pendugaan langsung dan pendugaan tidak langsung pada penelitian ini menggunakan pendekatan MSE *jackknife*. Didapatkan hasil perbandingan MSE antar kedua metode sesuai dengan tabel 4.5 dan tabel 4.6 yaitu pendugaan langsung menghasilkan rata-rata MSE sebesar 234 lebih besar dibandingkan dengan pendugaan tidak langsung dengan rata-rata MSE sebesar 211.03, perbandingan metode tersebut juga dapat dilihat secara grafik yang terdapat pada 4.10, kemudian dengan menggunakan metode RMSE yang merupakan akar dari MSE jackknife dimana hasil untuk RMSE kedua metode ini terdapat pada gambar 4.11 terlihat bahwa dengan menggunakan metode pendugaan tidak langsung menghasilkan nilai yang lebih kecil dibandingkan dengan pendugaan langsung.

# BAB V

# KESIMPULAN DAN SARAN

## 5.1 Kesimpulan

Berdasarkan analisis dan pembahasan yang telah diuraikan, dapat ditarik kesimpulan sebagai berikut:

1. Prosedur mendapatkan estimasi parameter model *small area estimation* berbasis area dengan pendekatan semiparametrik *penalized* spline, yang perlu diperhatikan bahwa model yang terbentuk merupakan hasil dari spline yang dilakukan penambahan nilai penalty untuk menghindari *overfitting* pada knot. Kemudian untuk mendapatkan model semiparametrik *small area estimation*, model *penalized spline* dikombinasikan dengan model *small area estimation* berbasis area agar dapat mendapatkan estimasi pendugaan area kecil secara semiparametrik dan menghasilkan nilai estimasi parameter dari model adalah $\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{Y}$ .
2. Pemodelan untuk pengeluaran per kapita dengan menggunakan pendekatan semiparametrik *penalized* spline di Kabupaten Sleman menghasilkan pengeluaran per kapita tertinggi di Kabupaten Sleman tahun 2014 yaitu terdapat pada kecamatan Ngaglik sebesar Rp 1.399.921,13, dengan menggunakan MSE *jackknife* untuk kebaikan model didapatkan hasil dengan menggunakan pendugaan tidak langsung menghasilkan nilai lebih kecil untuk tiap kecamatan, walaupun ada beberapa kecamatan yang memiliki hasil yang MSE yang nilainya cukup besar, hal ini dipengaruhi oleh pengeluaran per kapita di kecamatan tersebut yang memiliki hasil yang besar sehingga mempengaruhi MSE yang dihitung dengan menggunakan *jackknife.*

## 5.2 Saran

Saran yang bisa diberikan berdasarkan dari penelitian ini adalah sebagai berikut:

1. Pemerintah Kabupaten Sleman dapat melakukan pemerataan penduduk terhadap kecamatan yang mempunyai pengeluaran per kapita tinggi agar tidak terciptanya ketimpangan tiap daerah dan pemerintah Kabupaten Sleman juga lebih memperhatikan daerah yang letaknya jauh dari pusat kota
2. Penelitian ini merupakan penelitian yang menggunakan *small area estimation* berbasis area dengan observasi level kecamatan. Oleh karena itu untuk penelitian selanjutnya disarankan untuk melakukan *small area estimation* berbasis unit untuk mengetahui pengeluaran per kapita pada level desa dan agar dapat melihat perbandingan pengeluaran per kapita antar kecamatan dan desa.
3. Untuk penelitian selanjutnya juga menggunakan pendekatan *spatial empirical best liniear predictio*n (S-BLUP) agar dapat mengetahui deskripsi daerah yang diduga menggunakan pemetaan.

# DAFTAR PUSTAKA


Akita, T., dan Pirmansyah, L., (2011). *Urban Inequality in Indonesia*. IUJ Research Institute. Economics & Management Series. EMS 04.

Baskara, (2014). *Pendugaan Area Kecil Menggunakan Pendekatan Penalized Spline pada Pendugaan Pengeluaran Perkapita Tingkat Kecamatan Kabupaten Sumenep*. Tesis. Institut Teknologi Sepuluh November. Surabaya.

Badan Pusat Statistika. (2008). *DIY dalam Angka 2008*. Yogyakarta. BPS.

Badan Pusat Statistika. (2010). *Jawa Tengah Dalam Angka 2010.* Jawa Tengah. BPS

Badan Pusat Statistika. (2012). *DIY dalam Angka 2010*. Yogyakarta. BPS

Baillo, A. dan Molina, I. (2009). Mean-square errors of Small Area Estimators Under a Unit-level Multivariate Model, Statistics, 43, 553-569

Bapeda Sleman. (2012). *Indeks Pembangunan Manusia Kabupaten Sleman Tahun 2012*. http://bappeda.slemankab.go.id/indeks-pembangunan-manusia-kabupaten-sleman-tahun-2012.slm. Diakses 17 Agustus 2016.

Chandra, H., Chamber, R., dan Salvati, N. (2009). *Small Area Estimation of Proportion in Business Surveys*. Australia: The University of Wollongong.

Darsyah, M.Y. (2013). *Small Area Estimation terhadap Pengeluaran Per Kapita di Kabupaten Sumenep dengan Pendekatan Kernel-Boostrap*. Tesis (Tidak Dipublikasikan). Institut Teknologi Sepuluh Nopember. Surabaya.

Djuraidah, A., dan Aunuddin. (2006). *"Pendugaan Regresi Spline Terpenalti dengan Pendekatan Model Linear Campuran*. Statistika. Vol.6, No.1, hal.47-54.

Eubank, R. L. (1988). *Spline Smoothing and Nonparametric Regression*. New York: Marcel Deker.

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

Fathurahman, M. (2011). *Ëstimasi Parameter Model Regresi Spline*. Jurnal Eksponensial. Vol.2, No.1, hal.53-58.

Fausi, Hasan. (2011). *Small Area Estimation terhadap Pengeluaran Per kapita di Kabupaten Sumenep dengan menggunakan Metode Empirical Bayes*. Skripsi. Institut Teknologi Sepuluh Nopember. Surabaya

Griggs, W. (2013). *Penalized Spline Regression and Its Applications*. Senior Project. Whitman College. United States.

Giusti, C., Pratesi, M., dan Salvati, N., (2012). *A Semiparametric Fay-Herriot model using penalized spline*. Journal of The Indian Society Of Agricultural Statistics.

Gosh, dan Rao, J. (1994). Small Area Estimation: an Appraisal. *Statistical Science*. Vol. 9, 55-76.

Hall, P., dan Opsomer, J. (2005). *Theory for penalized spline regression*. Biomedika

Hardle, W. (1994). *Applied Nonparametric Regression*. NY: Cambrige University Press.

Krivoboka, T. (2006). Theoritical and Practical Aspects of Penalized Spline Smoothing. Disertasi. Universitas Bielefeld. Bielefeld

Kurnia, A., dan Notodiputro, K.A. (2006). *Penggunaan Metode Jackknife dalam Pendugaan Area Kecil*. Makalah disampaikan pada Seminar Nasional.

Mayasari. (2008). *Penerapan Metode Pemulusan Kernel pada Pendugaan Area Kecil*. SEMNAS Matematika dan Pendidikan Matematika.

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

Opsomer, J. (2004). *Penalized spline and small area estimation*. lowa State University.

Opsomer, J.D., Claeskens. G., Rannali, M.G., Kauermann. G., dan Breidt. F.J. (2008). *Nonparametric Small Area Estimation Using Penalized Spline Regression*. Journal of the Royal Statistical Society, Series B 70:265-286

Prasad, N.G.N., dan Rao, J.N.K. (1990). *The Estimation of The Mean Squared Error of The Small Area Estimators*. Journal of American Statistical Association, 85, hal.163-171.

Pratesi, M., Ranalli, M.G., dan Salvati, N. (2008). *Semiparametric M-Quantile regression for estimating the proportion of acidic lakes in 8-digit HUCs of the Northeastern US*. Environmetrics, Vol. 19, hal 659-764.

Rumiati, A.T. (2012). *Model Bayes untuk Pendugaan area kecil dengan penaksiran contoh berpeluang tidak sama pada kasus respon binomial dan multinomial*. Disertasi, Institut Pertanian Bogor. Bogor.

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

Ruppert, N.A. (2014). *Selecting the Number of Knots for Penalized Splines*. Journal of Computation and Graphical Statistics. Vol. 11 No.4, hal. 735-757.

Ruppert, D., Wand, M.P., dan Carrol, R.J. (2003), *Semiparametric Regression,* Cambridge University Press, New York.

Salam, N. (2013). *Estimasi Likelihood Maximum Penalized dari Model Regresi Semiparametrik*. Prosiding SEMNAS Universitas Diponegoro. Semarang.

Satriya, A. (2016). *Small Area Estimation Pengeluaran Per Kapita di Kabupaten Bangkalan dengan Metode Hierarchical Bayes*. Tesis (Tidak Dipublikasikan). Institut Teknologi Sepuluh Nopember. Surabaya.

Scealy. (2010). *Small Area Estimation Using a Multinomial Logit Mixed Model with Category Spesific Random Effects*. Austalian Bureau of Statistics and Australian National Univesity. Canberra: Australian Bureau of Statistics

Slemankab. (2013). *Kesadaran Masyarakat Sleman Ber-KB Semakin Tinggi*. http://www.slemankab.go.id/5071/kesadaran-masyarakat-sleman-ber-kb-makin-tinggi.slm. Diakses 28 Desember 2016

Tripena, A. (2011). *Penentuan Model Regresi Spline Terbaik". Prosiding Sewindu Statistika*. Universitas Diponegoro. Semarang. hal.92-102.

Tribun Jogja. (2015). *Kecamatan Depok Kawasan Paling Kumuh di Sleman.* http://jogja.tribunnews.com/2015/03/02/kecamatan-depok-kawasan-paling-kumuh-di-sleman. Diakses 4 Januari 2017

Wibowo, W. (2009). *Metode Kuadrat Terkecil Untuk Estimasi Kurva Regresi Semiparametrik Spline*. SEMNAS Matematika dan Pendidikan Matematika. FMIPA UNY. Yogyakarta.

Wulansari, I. (2013). *Pendugaan Area Kecil Terhadap Proporsi Rumah Tangga Miskin Level Kelurahan Di Kabupaten Sampang Menggunakan Hierarchical Bayes (HB) Logit Normal*. Semnas UNPAD.

Yao, F., dan Lee, T.C.M. (2008). *Ön Knot Placement for Penalized Spline Regression*. Journal of the Korean Statistical Society. No. 37, hal. 259-267.

*(Halaman ini sengaja dikosongkan)*

## LAMPIRAN 1. DATA PENELITIAN

### Data Pengeluaran Per Kapita Tingkat Kecamatan di Kabupaten Sleman berdasarkan Data Susenas 2014

| No | Kecamatan | Jumlah RT | Rata-rata Pengeluaran Per Kapita( $\times$ Rp 100.000) |
|---|---|---|---|
| 1 | Berbah | 30 | 4.32268724 |
| 2 | Cangkringan | 30 | 7.67818870 |
| 3 | Depok | 168 | 12.4765723 |
| 4 | Gamping | 59 | 11.0518897 |
| 5 | Godean | 57 | 6.22250643 |
| 6 | Kalasan | 39 | 11.7217523 |
| 7 | Minggir | 26 | 5.20199847 |
| 8 | Mlati | 74 | 11.2663689 |
| 9 | Moyudan | 20 | 7.0958158 |
| 10 | Ngaglik | 78 | 12.915154 |
| 11 | Ngemplak | 57 | 10.808763 |
| 12 | Pakem | 20 | 7.479988 |
| 13 | Prambanan | 28 | 5.7854442 |
| 14 | Seyegan | 30 | 6.76638752 |
| 15 | Sleman | 39 | 6.239480 |
| 16 | Tempel | 50 | 9.437896 |
| 17 | Turi | 10 | 7.503115 |

**LAMPIRAN 2. Data Variabel Penyerta Tingkat Kecamatan di Kabupaten Sleman**

| No | Kecamatan | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ |
|---|---|---|---|---|---|---|
| 1 | Berbah | 4798 | 14718 | 6586 | 3.5 | 1125 |
| 2 | Cangkringan | 2724 | 9740 | 4953 | 3.1 | 1076 |
| 3 | Depok | 10814 | 40956 | 23055 | 2.9 | 5244 |
| 4 | Gamping | 8361 | 27266 | 11671 | 3.4 | 2627 |
| 5 | Godean | 5879 | 19831 | 10403 | 3.1 | 3608 |
| 6 | Kalasan | 7271 | 24909 | 11888 | 3.1 | 3867 |
| 7 | Minggir | 2686 | 10748 | 4498 | 3.2 | 1765 |
| 8 | Mlati | 8145 | 27717 | 10221 | 2.8 | 2443 |
| 9 | Moyudan | 2568 | 11280 | 5401 | 2.7 | 1176 |
| 10 | Ngaglik | 8584 | 24622 | 10355 | 2.9 | 2326 |
| 11 | Ngemplak | 5399 | 16486 | 7138 | 3.5 | 1785 |
| 12 | Pakem | 3323 | 11574 | 7432 | 2.9 | 2950 |
| 13 | Prambanan | 4564 | 16606 | 9982 | 2.9 | 2122 |
| 14 | Seyegan | 4215 | 14288 | 6688 | 3.2 | 1567 |
| 15 | Sleman | 6390 | 20144 | 12350 | 3.3 | 797 |
| 16 | Tempel | 4856 | 16423 | 8034 | 3.1 | 850 |
| 17 | Turi | 3238 | 11351 | 5350 | 3.1 | 612 |

**Keterangan:**

$x_1$ : Keluarga Petani

$x_2$ : Keluarga Pengguna Listrik PLN

$x_3$: Jumlah Penduduk yang sedang sekolah (SD-SMK)

$x_4$ : Rata-rata anggota Keluarga

$x_5$: Kepadatan Penduduk

## LAMPIRAN 3. Hasil Dugaan dan Nilai Error untuk Variabel Parametrik

| No | Nilai dugaan parametrik | Error |
|---|---|---|
| 1 | 7.147054 | -2.82437 |
| 2 | 6.177324 | 1.500865 |
| 3 | 11.79028 | 0.686291 |
| 4 | 10.69037 | 0.361523 |
| 5 | 8.476369 | -2.25386 |
| 6 | 10.06603 | 1.65572 |
| 7 | 6.32975 | -1.12775 |
| 8 | 12.87856 | -1.61219 |
| 9 | 7.035417 | 0.060398 |
| 10 | 12.49841 | 0.416748 |
| 11 | 7.838926 | 2.969837 |
| 12 | 6.39441 | 1.085579 |
| 13 | 7.307646 | -1.5222 |
| 14 | 7.162439 | -0.39856 |
| 15 | 7.53914 | -1.29966 |
| 16 | 7.834577 | 1.603319 |
| 17 | 6.769351 | 0.733764 |

**LAMPIRAN 4. Nilai Estimasi Efek Acak Penalized Spline dengan 4 titik knot**

| Kecamatan | $\gamma_1$ | $\gamma_2$ | $\gamma_3$ | $\gamma_4$ | $\hat{v}$ |
|---|---|---|---|---|---|
| **Berbah** | -7.93E-14 | 7.83E-13 | 1.74E-12 | 2.66E-12 | -2.826386 |
| **Cangkringan** | -2.83E-12 | -2.48E-12 | -2.10E-12 | -1.74E-12 | -1.129504 |
| **Depok** | -9.56E-13 | -4.91E-13 | 2.67E-14 | 5.19E-13 | -1.524275 |
| **Gamping** | -3.43E-12 | -2.74E-12 | -1.98E-12 | -1.25E-12 | -2.256055 |
| **Godean** | 2.20E-13 | 6.17E-13 | 1.06E-12 | 1.48E-12 | -1.302124 |
| **Kalasan** | -1.18E-13 | 4.54E-15 | 1.41E-13 | 2.70E-13 | -0.4004947 |
| **Minggir** | 3.46E-15 | -1.45E-14 | -3.45E-14 | -5.35E-14 | 0.05888107 |
| **Mlati** | 1.22E-12 | 8.89E-13 | 5.21E-13 | 1.70E-13 | 1.083917 |
| **Moyudan** | -2.05E-13 | -4.29E-13 | -6.77E-13 | -9.14E-13 | 0.7321354 |
| **Ngaglik** | -2.10E-15 | -4.60E-13 | -9.69E-13 | -1.45E-12 | 1.49928 |
| **Ngemplak** | -2.20E-13 | -7.08E-13 | -1.25E-12 | -1.77E-12 | 1.601342 |
| **Pakem** | 1.26E-12 | 3.55E-13 | -6.53E-13 | -1.61E-12 | 2.967789 |
| **Prambanan** | 3.34E-13 | 2.25E-13 | 1.03E-13 | -1.33E-14 | 0.3588745 |
| **Seyegan** | -1.32E-12 | -8.32E-13 | -2.83E-13 | 2.38E-13 | -1.614183 |
| **Sleman** | 2.77E-12 | 2.27E-12 | 1.71E-12 | 1.17E-12 | 1.653285 |
| **Tempel** | 2.82E-13 | 7.35E-14 | -1.58E-13 | -3.79E-13 | 0.6822966 |
| **Turi** | 3.11E-13 | 1.85E-13 | 4.36E-14 | -9.05E-14 | 0.4152219 |

**LAMPIRAN 5. RMSE Pengeluaran Per Kapita dengan Menggunakan Metode Jackknife**

| Kecamatan | Rmse Penduga Langsung | Rmse Penduga Tidak Langsung |
|---|---|---|
| Berbah | 4.590799 | 4.470568 |
| Cangkringan | 1.501749 | 3.303059 |
| Depok | 1.559731 | 1.21632 |
| Gamping | 1.369892 | 1.157136 |
| Godean | 2.291637 | 1.633855 |
| Kalasan | 1.456429 | 1.115659 |
| Minggir | 3.290009 | 2.081403 |
| Mlati | 1.396628 | 1.707418 |
| Moyudan | 1.743291 | 1.363019 |
| Ngaglik | 1.619942 | 1.729077 |
| Ngemplak | 1.341225 | 1.221217 |
| Pakem | 1.573758 | 1.200496 |
| Prambanan | 2.664554 | 1.401237 |
| Seyegan | 1.923593 | 2.827826 |
| Sleman | 2.278596 | 1.121188 |
| Tempel | 1.244479 | 1.269082 |
| Turi | 1.564828 | 1.663288 |

## LAMPIRAN 6. Program Mencari Nilai Parametrik

```
parame<-read.csv("d://datafarida.csv",sep=";")
parame
y=as.vector(regi[,1])
X=as.matrix(cbind(1,regi[,2:5]))
y
X
p<- (dim(X)[2])-1#derajat bebas
b<- solve(t(X)%*%X)%*%t(X)%*%y# Estimator parameter model
y.hat<- X%*%b# Nilai dugaan
y.hat
error<-y-y.hat
error
```

**LAMPIRAN 7. Program Pencarian Model Regresi Nonparametrik Menggunakan Penalized Spline**

```
 datafix<-matrix(scan("D://run data tesis.txt"), byrow=T,17,2)
 datafix
 K<- min(0.25*length(unique(x)),35)
 d<- 1
 k <- K+d
 knot<- quantile(unique(x),probs=(1:(k-d))/(k-d+1),names =
FALSE)
 X<- 1
 for (p in 1:d) X <-cbind(X,x^p)
 for (j in 1:K) X <-cbind(X,(pmax(0,x-knot[j]))^d)
 X
 D<- diag(2+K)
 D[1,1]<- 0
 D[2,2]<- 0
D[3,3]<- 0
 In<- diag(n)
 GCV2<-function(lamda)
 { A <- X%*%ginv(t(X)%*%X+lamda*D)%*%t(X)
 IA <- In-A
 GCV2 <-(1/n)*t(IA%*%y)%*%(IA%*%y)/((sum(diag(IA))/n)^2)
 return(GCV2)
 }
 opt2 <- nlminb(c(0),GCV2,lower=c(0),upper=c(Inf)) #iterasi
untuk mencari parameter optimum
 l.opt <- opt2$par
 GCV.opt<- opt2$objective
 b<- ginv(t(X)%*%X)%*%t(X)%*%y
 y.hat<- X%*%b
 SST<- sum((y-mean(y))^2)
SSR<- sum((y.hat-mean(y))^2)
 SSE<- sum((y-y.hat)^2)
 r<- dim(X)[2]-1
 MSR<- SSR/r
 MSE<- SSE/(n-r-1)
 R2<- SSR/SST
 cov.b<- ginv(t(X)%*%X)*MSE
 se.b<- sqrt(diag(cov.b))
 t.hit<- (1/se.b)*b
 p.val<- 2*(1-pt(abs(t.hit),n-2))
 res<- y-y.hat
 hasil<- cbind(b,se.b,t.hit,p.val)
plot(x,y)
lines(x[order(x)],y.hat[order(x)],col="red")
lines(rep(1078,17),y)
lines(rep(1585,17),y)
lines(rep(2151,17),y)
lines(rep(2688,17),y)
```

**Lampiran 8. Program Pendugaan Area Estimation dengan Pendekatan Semiparametrik**

```
data=read.csv("d:/ =";")  #menginput data
y=as.vector(data[,1])
X=as.matrix(cbind(1,data[,2:5]))
knot=c(1078.33,1585.85,2151.14,2688.52)
z.spline=outer(data[,6],knot,"-")
idnum=1:17
id=factor(idnum)
group=1:17
n=nrow(X)
m=n
ZBlock=pdBlocked(list(pdIdent(~z.spline-1),pdIdent(~id-1)))
dataFr=groupedData(y~x1+x2+x3+x4|group,data=data.frame(y,X,z.spl
ine,idnum))
fit=lme(fixed=y~-1+X,data=dataFr, random=ZBlock)
beta.hat=as.matrix(fit$coef$fixed)
yhat=X%*%beta.hat
random=fit$coef$random$group[,1:4]
write.csv(random,"a.csv")
rand=read.csv("a.csv")
ran.sr=rand[order(rand[,1]),]
random.sort=ran.sr[,2:ncol(ran.sr)]
sum.rand.sort=NULL
for (i in 1:n)
{ sum.rand.sort[i]=sum(random.sort[i,])}
tetahat=yhat+sum.rand.sort
me=mean(tetahat)
sde=sd(tetahat)
teta.i=tetahat+u
```

## Lampiran 9. Program MSE Jackknife

```
library(lmeSplines)
library(nlme)
data=read.csv("d:/datafarida.csv",sep=";")  #menginput data
y=as.vector(data[,1])
X=as.matrix(cbind(1,data[,2:5]))
knot=c(1078.33,1585.85,2151.14,2688.52)
z.spline=outer(data[,6],knot,"-")
idnum=1:17
id=factor(idnum)
group=1:17
n=nrow(X)
m=n
ZBlock=pdBlocked(list(pdIdent(~z.spline-1),pdIdent(~id-1)))
dataFr=groupedData(y~x1+x2+x3+x4|group,data=data.frame(y,X,z.spl
ine,idnum))
fit=lme(fixed=y~-1+X,data=dataFr, random=ZBlock)
beta.hat=as.matrix(fit$coef$fixed)
yhat=X%*%beta.hat
random=fit$coef$random$group
write.csv(random,"a.csv")
rand=read.csv("a.csv")
ran.sr=rand[order(rand[,1]),]
random.sort=ran.sr[,2:ncol(ran.sr)]
z.random=random.sort[,1:4]
id.random=random.sort[,5:ncol(random.sort)]
tetahat=NULL
error=NULL
for (i in 1:n)
{
  tetahat[i]=yhat[i]+sum(z.random[i,])+id.random[i,i]
  error[i]=y[i]-id.random[i,i]
}
#===========
se=sd(error)^2
y.ybar.2=(y-mean(y))^2
sv=(1/(m-1)*sum(y.ybar.2))-se
me2=NULL
sv.u=NULL
y.me=NULL
svu.sv=NULL
h1=NULL
for (i in 1:n)
{
  y2=y[-i]
  me2[i]=mean(y2)
  y.me[i]=(y[i]-me2[i])^2
  sv.u[i]=(1/(m-2)*y.me[i])-se
  svu.sv[i]=((m-1)/(m-2)*sv.u[i])-sv
  h1[i]=sv-svu.sv[i]
}
#==================================
tetahat2=matrix(0,m-1)
idnum2=idnum[-n]
id2=factor(idnum2)
group2=group[-n]
```

```
tetahat2=matrix(0,n-1,n)
J=matrix(0,n-1,n)
h2=NULL
for (i in 1:m)
{
  data=read.csv("d:/datafarida.csv",sep=";")
  write.csv(data[-i,],"b.csv")
  data2=read.csv("b.csv")
  y=as.vector(data2[,2])
  X=as.matrix(cbind(1,data2[,3:6]))
  z.spline=outer(data2[,7],knot,"-")
  ZBlock=pdBlocked(list(pdIdent(~z.spline-1),pdIdent(~id2-1)))

dataFr=groupedData(y~x1+x2+x3+x4|group2,data=data.frame(y,X,z.sp
line,idnum2))
  fit2=lme(fixed=y~-1+X,data=dataFr, random=ZBlock)
  beta.hat2=as.matrix(fit2$coef$fixed)
  yhat2=X%*%beta.hat2
  random2=fit2$coef$random$group
  write.csv(random2,"a.csv")
  rand2=read.csv("a.csv")
  ran.sr2=rand2[order(rand2[,1]),]
  z.random=ran.sr2[,2:5]
  id.random=ran.sr2[,6:ncol(random.sort)]
  for (j in 1:(n-1))
  {
    tetahat2[j,i]=yhat2[j]+sum(z.random[j,])+id.random[j,j]
    J[j,i]=(tetahat2[j,i]-tetahat[i])^2
  }
  h2[i]=((m-1)/m)*sum(J[,i])
}
MSE=h1+h2
```

## BIODATA PENULIS

Farida Apriani lahir pada 19 April 1992 di Pangkalpinang, Bangka Belitung sebagai anak Bungsu dari lima bersaudara. Jenjang pendidikan yang telah ditempuh Sekolah Dasar SD Negeri 3 Pangkalpinang pada tahun 1998-2004, kemudian pendidikan menengah pertama ditempuh di SMP Negeri 2 Pangkalpinang pada tahun 2004-2007. Pada tahun 2007-2010 melanjutkan pendidikan menengah atas di SMA Tunas Harapan Bangsa . Pendidikan tinggi dimulai pada tahun 2010 di Program Studi Matematika dan Ilmu Pengetahuan Alam (FMIPA), Jurusan Statistika, Universitas Islam Indonesia, Yogyakarta dan menyelsaikan program S-1 pada tahun 2014. Kemudian tahun 2015 melanjutkan program pendidikan S-2 di Institut Sepuluh Nopember (ITS), Jurusan Statistika, Fakultas Matematika dan Imu Pengetahuan Alam (FMIPA). Jika ada saran dan kritik mengenai tugas akhir yang penulis buat ini dapat menghubungi penulis melalui Faridaprn36@gmail.com