SKRIPSI

PEMODELAN ANALISIS KOMPONEN UTAMA REGRESI TERBOBOTI SECARA GEOGRAFIS UNTUK DATA YANG MENGALAMI MULTIKOLINEARITAS

Disusun dan diajukan oleh

MUSDALIFAH

H051171010

PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

MAKASSAR

2021

PEMODELAN ANALISIS KOMPONEN UTAMA REGRESI TERBOBOTI SECARA GEOGRAFIS UNTUK DATA YANG MENGALAMI MULTIKOLINEARITAS

SKRIPSI

Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Sains Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin Makassar

MUSDALIFAH

H051171010

PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

MAKASSAR

2021

LEMBAR PERNYATAAN KEASLIAN

Yang bertanda tangan dibawah ini :

Nama	: Musdalifah
NIM	: H051171010
Program Studi	: Statistika
Jenjang	: Sarjana (S1)

Menyatakan dengan ini bahwa karya tulisan saya berjudul

PEMODELAN ANALISIS KOMPONEN UTAMA REGRESI TERBOBOTI SECARA GEOGRAFIS UNTUK DATA YANG MENGALAMI MULTIKOLINEARITAS

Adalah karya tulisan saya sendiri dan bukan merupakan pengambil alihan tulisan orang lain bahwa skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri.

Apabila dikemudian hari terbukti atau dapat dibuktikan bahwa sebagian atau keseluruhan skripsi ini hasil karya orang lain, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Makassar, 22 Februari 2021

METERAI TEMPEL
TGL. 20
320D1AHF892604594
6000
ENAM RIBU RUPIAH

MUSDALIFAH

NIM. H051171010

PEMODELAN ANALISIS KOMPONEN UTAMA REGRESI TERBOBOTI SECARA GEOGRAFIS UNTUK DATA YANG MENGALAMI MULTIKOLINEARITAS

Disetujui Oleh:

Pembimbing Utama,	**Pembimbing Pertama,**
Andi Kresna Jaya, S.Si., M.Si.	**Dra. Nasrah Sirajang, M.Si.**
NIP. 19731228 200003 1001	**NIP. 19650519 1999303 2002**

Ketua Departemen Statistika

Dr. Nurtiti Sunusi, S.Si., M.Si.
NIP. 19720117 199703 2002

Pada Tanggal : 22 Februari 2021

LEMBAR PENGESAHAN (TUGAS AKHIR)

PEMODELAN ANALISIS KOMPONEN UTAMA REGRESI TERBOBOTI SECARA GEOGRAFIS UNTUK DATA YANG MENGALAMI MULTIKOLINEARITAS

Disusun dan diajukan oleh

MUSDALIFAH

H051171010

Telah dipertahankan di hadapan Panitia Ujian yang dibentuk dalam rangka Penyelesaian Studi Program Sarjana Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin pada tanggal 22 Februari 2021 dan dinyatakan telah memenuhi syarat kelulusan.

Menyetujui,

Pembimbing Utama,	**Pembimbing Pertama,**
Andi Kresna Jaya, S.Si., M.Si.	**Dra. Nasrah Sirajang, M.Si.**
NIP. 19731228 200003 1001	**NIP. 19650519 1999303 2002**

Ketua Departemen Statistika

Dr. Nurtiti Sunusi, S.Si., M.Si.

NIP. 19720117 199703 2002

KATA PENGANTAR

Assalamu'alaikum Warohmatullahi Wabarokatuh.

Alhamdulillahi robbil'alamin, Puji syukur kepada Allah SWT atas segala limpahan rahmat, nikmat, dan hidayah yang diberikan kepada penulis sehingga dapat menyelesaikan penulisan skripsi dengan judul **"Pemodelan Analisis Komponen Utama Regresi Terboboti Secara Geografis Untuk Data yang Mengalami Multikolinearitas"** sebagai salah satu syarat untuk memperoleh gelar Sarjana Sains pada Program Studi Statistika Jurusan Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam.

Salam dan sholawat *InsyaAllah* senantiasa tercurah kepada **Nabi Muhammad *Shallallahu'alaihi Wasallam***, sang kekasih tercinta yang telah memberikan petunjuk cinta dan kebenaran dalam kehidupan.

Dalam penyelesaian skripsi ini, penulis telah melewati perjuangan panjang dan pengorbanan yang tidak sedikit. Namun berkat rahmat dan izin-Nya serta dukungan dari berbagai pihak yang turut membantu baik moril maupun material sehingga akhirnya tugas akhir ini dapat terselesaikan. Oleh karena itu, penulis menyampaikan ucapan terima kasih yang setinggi-tingginya dan penghargaan yang tak terhingga kepada Ayahanda **Iming** dan Ibunda tercinta **Hasriani** yang telah membesarkan dan mendidik penulis dengan penuh kesabaran dan dengan limpahan cinta dan kasih saying serta telah memberikan dukungan dan doa kepada penulis yang tak pernah habis.

Ucapan terima kasih yang juga penulis ucapkan kepada:

1. **Ibu Prof. Dr. Dwia Aries Tina Pulubuhu, MA,** selaku Rektor Universitas Hasanuddin beserta seluruh jajarannya.
2. **Bapak Dr. Eng. Amiruddin,** selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin beserta seluruh jajarannya.
3. **Ibu Dr. Nurtiti Sunusi, S.Si., M.Si.,** selaku Ketua Departemen Statistika yang telah seperti orang tua sendiri. Segenap dosen pengajar dan staf **Departemen**

Statistika yang telah membekali ilmu dan kemudahan kepada penulis dalam berbagai hal selama menjadi mahasiswa di Departemen Statistika,

4. **Bapak Andi Kresna Jaya, S.Si., M.Si.,** selaku Pembimbing Utama sekaligus penasehat akademik penulis yang telah meluangkan waktu dan pemikirannya untuk memberikan arahan, pengetahuan dan bimbingan kepada penulis mulai dari awal hingga selesainya penulisan tugas akhir ini. Di mata penulis beliau adalah seorang pendidik yang luar biasa terhadap mahasiswa di Departemen Statistika. **Ibu Dra. Nasrah Sirajang, M.Si.,** selaku Pembimbing Pertama penulis yang telah meluangkan waktunya ditengah kesibukan untuk memberikan arahan bagi penulis.
5. **Ibu Anisa, S.Si., M.Si.** dan **Ibu Sri Astuti Thamrin, S.Si., M.Stat., Ph.D.,** selaku tim penguji yang telah memberikan saran dan pertanyaan yang tak terhingga nilainya dalam penyusunan skripsi ini.
6. **Bapak Andi Masjaya, S.Pd., M.Pd.,** dan **Ibu Halija S.Pd.,** yang telah seperti orang tua penulis, menyayangi dan mengasihi penulis seperti anak sendiri dan yang selalu memberi motivasi penulis hingga dapat menyelesaikan tugas akhir ini. Penulis sangat berterima kasih atas kebaikan beliau, dan juga untuk **Andi Riski Mujahida Masdi, S.H.,** terima kasih telah menjadi segalanya bagi penulis.
7. Spesial untuk orang terdekat penulis, **Nurul Fitriani Arfanti** yang telah menjadi sahabat terbaik dan senantiasa memberikan dorongan, semangat, dan motivasi dalam setiap keadaan.
8. Sahabat seperjuangan yang selalu menemaniku dibangku kuliah, the real is "**BONE BARRU" Nur Kamalia, Fitri** dan **Nur Alya Tussa'ada**. Terima kasih atas kebersamaan, kebahagiaan, kesedihan, dan kebaikannya selama masa perkuliahan. Tetap semangat dan sukses sist!
9. Teman-teman seperjuangan "**LIM17"** yang senantiasa memberikan dukungan, semangat, dan motivasi dalam menyelesaikan tugas akhir ini.
10. Teman-teman **Statistika 2017**, terima kasih atas kebersamaan, suka dan duka selama menjalani pendidikan di Departemen Statistika.

11. **Kak Reski Amalah, Kak Suritman, Kak Samsul Arifin, Fatir, Kukuh Adi dan Riri** terima kasih atas segala bentuk bantuan selama penulis mengerjakan tugas akhir ini.
12. Teman-teman **KKN Bone 3 Gelombang 104**, terima kasih telah menjadi teman sekaligus keluarga selama sebulan lebih, semoga silaturahmi tetap terjalin.
13. Kepada seluruh pihak yang tidak dapat penulis sebutkan satu persatu, terima kasih setinggi-tingginya untuk segala dukungan dan partisipasi yang diberikan kepada penulis semoga bernilai ibadah di sisi **Allah SWT.**

Penulis berharap skripsi ini dapat memberikan tambahan pengetahuan baru bagi para pembelajar statistika. Penulis menyadari bahwa dalam penulisan tugas akhir ini masih banyak terdapat kekurangan. Oleh karena itu, dengan segala kerendahan hati penulis memohon maaf. Akhir kata, semoga dapat bermanfaat bagi pihak-pihak yang berkepentingan. *Aamiin Yaa Rabbal Alamin.*

Makassar, 22 Februari 2021

Musdalifah

PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

Sebagai civitas akademik Universitas Hasanuddin, saya yang bertandatangan di bawah ini:

Nama	: Musdalifah
NIM	: H051171010
Program Studi	: Statistika
Departemen	: Statistika
Fakultas	: Matematika dan Ilmu Pengetahuan Alam
Jenis Karya	: Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Hasanuddin **Hak Bebas Royalti Non eksklusif (*Non-exclusive Royalty- Free Right*)** atas tugas akhir saya yang berjudul:

"PEMODELAN ANALISIS KOMPONEN UTAMA REGRESI TERBOBOTI SECARA GEOGRAFIS UNTUK DATA YANG MENGALAMI MULTIKOLINEARITAS"

Beserta perangkat yang ada (jika diperlukan). Terkait dengan hal di atas, maka pihak universitas berhak menyimpan, mengalih-media/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan memublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di Makassar pada tanggal 22 Februari 2021

Yang menyatakan

Musdalifah

ABSTRAK

Analisis regresi linier adalah metode yang digunakan untuk memodelkan hubungan antara suatu variabel respon dengan variabel penjelas. Pada kenyataannya, antara lokasi satu dengan lainnya memiliki kondisi yang berbeda, sehingga data antar pengamatan sulit dianalisis menggunakan regresi linier. Keragaman spasial model regresi akan mengakibatkan hasil yang diperoleh kurang sesuai. Regresi Terboboti Geografis (RTG) merupakan metode statistika yang digunakan dalam menganalisis keragaman spasial. Dalam aplikasi RTG terkadang ditemukan adanya multikolinearitas yang dapat mengakibatkan parameter model regresi yang dihasilkan memiliki sisa yang sangat besar. Analisis Komponen Utama (AKU) merupakan metode yang digunakan untuk mengatasi masalah multikolinearitas. Penelitian ini bertujuan untuk mendapatkan estimasi parameter model analisis komponen utama regresi terboboti secara geografis pada data yang terdapat masalah multikolinearitas. Penelitian ini diaplikasikan pada data Produk Domestik Regional Bruto (PDRB) kabupaten/kota di Sulawesi Selatan tahun 2019. Hasil penelitian menunjukkan model analisis komponen utama regresi terboboti secara geografis lebih baik digunakan dalam memodelkan PDRB di Sulawesi Selatan tahun 2019 jika terdapat masalah multikolinearitas. Dalam analisis komponen utama regresi terboboti secara geografis variabel yang berpengaruh signifikan di 16 Kabupaten/Kota di Sulawesi Selatan diwakili dengan variabel baru (komponen utama) hasil analisis komponen utama (AKU) yaitu w_1, w_2 dan w_3 yang mampu menerangkan total varian data 85,07%.

Kata Kunci: PDRB, Multikolinearitas, Regresi Terboboti Geografis, Analisis Komponen Utama, Keragaman spasial.

ABSTRACT

Linear regression analysis is a method for modeling the relation between a response variable with independents variable. In fact, the location with others has different conditions, so the data between observations was difficult to be analyzed using linear regression. The spatial heterogeneity of the regression models will give less well-matched results. Geographical Weighted Regression (GWR) is a statistical method used in analyzing spatial heterogeneity. In the GWR application, multicollinearity is sometimes found which can result in the resulting parameter regression model having a very large residual. Principal Component Analysis (PCA) is a method used to solve multicollinearity problems. This study aims to obtain parameter estimation model of the principal component analysis of geographically weighted regression on data with multicollinearity problems. This study was applied to data on the Gross Regional Domestic Product (GRDP) of districts/cities in South Sulawesi in 2019. The showed that the principal component analysis model of geographically weighted is better to use in modeling the GRDP if South Sulawesi in 2019 there are multicollinearity problems. In principal component analysis, geographically weighted regression the variables that have a significant effect on 16 districts/cities in South Sulawesi are represented by new variables (principal components) from the results of the PCA, that is w_1, w_2 *and* w_3 *which are able to explain the total data variance of 85.07%.*

Keywords: GRDP, Multicollinearity, Geographically Weighted Regression, Principal Component Analysis, Spatial heterogeneity.

DAFTAR ISI

Halaman

DAFTAR LAMPIRAN

DAFTAR GAMBAR

DAFTAR TABEL

BAB I
PENDAHULUAN

1.1 Latar Belakang

Analisis regresi merupakan suatu metode statistika yang digunakan untuk menganalisis hubungan antara variabel respon dengan variabel penjelas yang umumnya dinyatakan dalam persamaan matematika. Apabila ada pengaruh unsur wilayah atau data spasial dalam model regresi, maka metode regresi linier akan kurang tepat dilakukan karena efek spasial tidak tercakup di dalamnya (Anselin, 1988). Data spasial memuat informasi tentang data yang dikumpulkan dari lokasi yang berbeda dan terdapat ketergantungan antara pengukuran data dengan lokasi. Sedangkan efek spasial terkait dengan perbedaan budaya, geografis, lingkungan dan kebijakan ekonomi yang bervariasi di setiap lokasi. Efek spasial ini kemudian disebut sebagai keragaman spasial atau heterogenitas spasial.

Heterogenitas spasial muncul jika residual dari model yang diamati memiliki varian yang tidak konstan dari suatu lokasi pengamatan ke lokasi pengamatan lainnya. Salah satu metode statistika untuk mengatasi masalah heterogenitas spasial adalah Regresi Terboboti Geografis (RTG), yaitu metode yang menggunakan faktor geografis sebagai variabel penjelas yang dapat mempengaruhi variabel respon (Fotheringham, 2002).

RTG merupakan pengembangan dari model regresi yang setiap parameter dihitung berdasarkan setiap lokasi pengamatan, sehingga setiap lokasi pengamatan mempunyai nilai pendugaan parameter regresi yang berbeda-beda atau bersifat lokal (Fotheringham, 2002). Pendugaan parameter pada model RTG membutuhkan matriks pembobot menggunakan metode *Weighted Least Square* (WLS). WLS adalah pendugaan parameter yang diberi pembobot. Matriks pembobot dibentuk menggunakan fungsi pembobot dengan *bandwidth*. Salah satu cara untuk mendapatkan nilai *bandwidth* yang berbeda pada setiap lokasi pengamatan yaitu dengan menggunakan fungsi pembobot *adaptive gaussian kernel.*

Dalam aplikasi RTG terkadang ditemukan adanya variabel penjelas yang saling berkorelasi, permasalahan tersebut dinamakan multikolinearitas lokal.

Multikolinearitas lokal adalah suatu kondisi dengan variabel penjelas saling berkorelasi di setiap lokasi pengamatan (Ramadhan, 2012). Adanya multikolinearitas dapat menyebabkan pendugaan parameter dari model yang dihasilkan memiliki variansi yang besar, sehingga dapat menyebabkan terjadinya kesalahan dalam menginterpretasikan parameter. Salah satu cara untuk mendeteksi adanya multikolinearitas lokal pada model spasial dengan melihat nilai dari Variance Inflation Factor (VIF). Nilai VIF menunjukkan terjadi multikolinearitas pada data apabila terdapat variabel yang memiliki nilai VIF>10 (Putri, 2011).

Pada regresi spasial, multikolinearitas dapat diatasi menggunakan konsep dari Analisis Komponen Utama (AKU). Penambahan AKU pada model RTG diharapkan mampu mengatasi masalah multikolinearitas lokal dengan lebih sederhana. AKU adalah metode analisis multivariat yang mentransformasi variabel-variabel asal yang saling berkorelasi menjadi variabel-variabel baru yang tidak saling berkorelasi dengan mereduksi sejumlah variabel tersebut sehingga mempunyai dimensi yang lebih kecil namun dapat menerangkan sebagian besar keragaman variabel aslinya (Johnson, 2002).

Banyaknya komponen utama yang terbentuk sama dengan banyaknya variabel asli. Pereduksian dimensi dilakukan dengan kriteria persentase keragaman data yang diterangkan oleh beberapa komponen utama pertama. Apabila beberapa komponen utama pertama telah menerangkan lebih dari 70% sampai 85% keragaman data asli, maka analisis cukup dilakukan sampai dengan komponen utama tersebut (Bilson, 2005).

Model RTG dapat diterapkan pada kasus Produk Domestik Regional Bruto (PDRB) kabupaten/kota di Sulawesi Selatan. Data PDRB dan faktor-faktor yang mempengaruhinya akan berbeda untuk setiap daerah atau lokasi pengamatan, tergantung pada kondisi daerah atau lokasi pengamatan yang diamati. Meskipun demikian kondisi suatu lokasi pengamatan akan memiliki hubungan yang erat dengan lokasi pengamatan yang berdekatan.

PDRB merupakan nilai tambah bruto seluruh barang dan jasa yang tercipta atau dihasilkan di wilayah domestik suatu negara yang timbul akibat berbagai aktivitas ekonomi dalam suatu periode tertentu. Penyusutan PDRB dapat dilakukan melalui tiga pendekatan yaitu pendekatan produksi, pengeluaran dan pendapatan yang disajikan atas dasar harga yang berlaku. PDRB harga berlaku bertujuan untuk melihat struktur perekonomian yang menunjukkan kemampuan sumber daya ekonomi yang dihasilkan oleh suatu daerah. Selain itu, kondisi perekonomian secara keseluruhan di setiap daerah juga dapat dilihat dari seberapa besar jumlah belanja daerah pada daerah bersangkutan. Belanja daerah merupakan bentuk rangsangan yang dilakukan oleh pemerintah terhadap perekonomian daerah. Jika nilai belanja daerah yang dialokasikan untuk pembangunan besar, maka akan meningkatkan kesejahteraan penduduk. Hal ini berarti kondisi ekonomi di daerah tersebut akan meningkat.

Badan Pusat Statistik (BPS) menyatakan bahwa jumlah perekonomian Sulawesi Selatan berdasarkan besaran PDRB atas dasar harga yang berlaku tahun 2019 mencapai 5,13 %. Nilai PDRB mengalami penurunan sebesar 0,22% jika dibandingkan nilai PDRB tahun 2018 dengan nilai pada periode tersebut sebesar 5,35 % (BPS, 2019). Sehubungan dengan data PDRB di Sulawesi Selatan yang telah diuraikan. Data yang digunakan pada penelitian ini yaitu data PDRB kabupaten/kota dan mengidentifikasi faktor-faktor yang berpengaruh terhadap Produk Domestik Regional Bruto (PDRB) kabupaten/kota di Provinsi Sulawesi Selatan, dengan memperhatikan faktor geografis.

Beberapa penelitian yang telah menggunakan metode RTG diantaranya Rahmawati dan Anik Djuraidah (2010) menggunakan RTG dengan pembobot kernel kuadrat ganda, Nufusia, dkk (2013) menggunakan RTG untuk menganalisis pendapatan asli daerah kabupaten/kota di Provinsi Aceh. Akan tetapi, dalam menganalisis model RTG terkadang ditemukan multikolinearitas lokal disebabkan adanya hubungan antar variabel bebas yang mengakibatkan nilai variansi pada pendugaan parameter besar. Oleh karena itu, penelitian penyelesaian multikolinearitas dikembangkan oleh Abdullah, dkk (2017) menggunakan RTG model regresi *ridge* yang melakukan transformasi pada masing-masing variabel melalui prosedur pemusatan dan penskalaan. Kemudian menambahkan koefisien tetapan bias positif pada proses pendugaan parameter. Sehingga berdasarkan latar belakang tersebut, peneliti tertarik untuk membahas pemodelan analisis komponen utama regresi terboboti secara geografis untuk data yang mengalami multikolinearitas.

1.2 Rumusan Masalah

Adapun rumusan masalah dalam penelitian ini sebagai berikut:

1. Bagaimana bentuk estimasi parameter analisis komponen utama regresi terboboti secara geografis untuk data yang mengalami multikolinearitas?
2. Bagaimana menentukan model analisis komponen utama regresi terboboti secara geografis untuk data yang mengalami multikolinearitas pada studi kasus Produk Domestik Regional Bruto (PDRB) kabupaten/kota di Sulawesi Selatan tahun 2019 ?

1.3 Batasan Masalah

Penelitian ini memiliki permasalahan yang dibatasi pada data yang mengalami heterogenitas spasial dan terdapat multikolinearitas, kemudian menentukan model RTG dengan analisis komponen utama menggunakan fungsi pembobot *adaptive gaussian kernel.*

1.4 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah sebagai berikut:

1. Mendapatkan estimasi parameter model analisis komponen utama regresi terboboti secara geografis untuk data yang mengalami multikolinearitas.
2. Mendapatkan model analisis komponen utama regresi terboboti secara geografis untuk data yang mengalami multikolinearitas pada studi kasus Produk Domestik Regional Bruto (PDRB) kabupaten/kota di Sulawesi Selatan tahun 2019.

1.5 Manfaat Penelitian

Hasil penelitian ini diharapkan dapat memberikan manfaat berupa tambahan kepustakaan bagi pengguna ilmu statistika tentang cara mengatasi data yang multikolinearitas dengan menggunakan metode Analisis Komponen Utama (AKU) dalam model Regresi Terboboti Geografis (RTG), dan pemahaman dalam analisis persentase Produk Domestik Regional Bruto (PDRB) kabupaten/kota di Sulawesi Selatan dengan faktor-faktor yang mempengaruhinya.

BAB II

TINJAUAN PUSTAKA

2.1 Model Regresi Linier Berganda

Analisis regresi linier berganda merupakan salah satu metode statistika yang digunakan untuk memodelkan hubungan antara sebuah variabel respon dengan dua atau lebih variabel penjelas (Montgomery, 2001). Secara matematis, model regresi linier berganda dapat dirumuskan sebagai berikut:

$$y_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} + \varepsilon_i, \quad i = 1, 2, \dots, n\,; j = 1,2, \dots , p$$

Model persamaan regresi linier berganda dapat ditulis dalam bentuk matriks, sebagai berikut:

$$\mathbf{Y} = \mathbf{X\beta} + \boldsymbol{\varepsilon} \tag{2.1}$$

$$\boldsymbol{Y} = (y_1 \;\; y_2 \;\; y_3 \dots y_n)', \boldsymbol{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & x_{13} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & x_{23} & \dots & x_{2p} \\ 1 & x_{31} & x_{32} & x_{33} & \dots & x_{3p} \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & x_{n1} & x_{n2} & x_{n3} & \dots & x_{np} \end{pmatrix},$$

$$\boldsymbol{\beta} = (\beta_0 \;\; \beta_1 \;\; \beta_2 \;\dots\; \beta_p)', \boldsymbol{\varepsilon} = (\varepsilon_1 \;\; \varepsilon_2 \;\; \varepsilon_3 \;\dots\; \varepsilon_n)'$$

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & x_{13} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & x_{23} & \dots & x_{2p} \\ 1 & x_{31} & x_{32} & x_{33} & \dots & x_{3p} \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & x_{n1} & x_{n2} & x_{n3} & \dots & x_{np} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

$$\boldsymbol{Y} = n \times 1 \;\;;\;\; \boldsymbol{X} = n \times (p+1) \;\;;\;\; \boldsymbol{\beta} = (p+1) \times 1 \;\;;\;\; \boldsymbol{\varepsilon} = n \times 1$$

Estimasi parameter menggunakan metode OLS diperoleh:

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y} \tag{2.2}$$

2.2 Multikolinearitas

Multikolinearitas adalah suatu kondisi terjadi korelasi antara satu variabel bebas dengan variabel bebas yang lainnya sehingga menyebabkan parameter dalam model regresi yang dihasilkan memiliki sisa yang sangat besar. Dalam pemodelan RTG, perhitungan nilai VIF dilakukan untuk masing-masing variabel penjelas. Nilai VIF menunjukkan terjadi multikolinearitas pada data apabila terdapat variabel yang memiliki nilai VIF>10 (Putri, 2011). Nilai VIF dinyatakan sebagai berikut:

$$\text{VIF} = \frac{1}{1-R_j^2} \tag{2.3}$$

dengan R_j^2 merupakan nilai koefisien determinasi antara x_j dengan variabel prediktor lainnya pada persamaan regresi. Apabila nilai VIF > 10, maka dapat diindikasikan bahwa terdapat kasus multikolinieritas yang serius

2.3 Pengujian Pengaruh Spasial

Pengaruh spasial yang berkaitan dengan pendekatan titik dan perbedaan karakteristik lingkungan serta geografis antar wilayah pengamatan dapat menyebabkan keragaman variansi atau heterogenitas spasial (Anselin, 2011). Heterogenitas spasial adalah apabila satu variabel penjelas yang sama memberikan respon yang tidak sama pada lokasi yang berbeda dalam satu lokasi pengamatan.

Pengujian heterogenitas spasial ini dapat dilakukan dengan menggunakan statistik uji *Breusch-Pagan* (BP). Bentuk hipotesis pengujian:

$H_0 : {\sigma_1}^2 = {\sigma_2}^2 = \cdots = {\sigma_n}^2 = \sigma^2$

(Tidak terdapat heterogenitas spasial)

H_1 : ada ${\sigma_i}^2 \neq \sigma^2$ $, i = 1,2, \dots, n$

(Ada heterogenitas spasial)

Statistik Uji:

$$BP = \left(\frac{1}{2}\right) \boldsymbol{f}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{f} \tag{2.4}$$

dengan elemen vektor $\boldsymbol{f}$ adalah $f_i = \left(\frac{{e_i}^2}{\sigma^2} - 1\right)$, e_i = galat kuadrat terkecil untuk observasi ke-i, σ^2 = ragam galat e_i, dan $\boldsymbol{Z}$ = matriks berukuran $n \times (p+1)$ berisi vektor yang telah terstandarisasi (**z**) untuk setiap lokasi dan *p* merupakan banyaknya variabel prediktor. Kriteria keputusan adalah tolak H_0 jika $BP > \chi^2_{(\alpha,p)}$ sehingga dapat disimpulkan terdapat heterogenitas spasial (Anselin, 2009 dalam Ramadhan, 2012).

2.4 Uji Asumsi Komponen Utama

Dalam penggunaan Analisis Komponen Utama (AKU), terdapat asumsi yang harus dipenuhi. Asumsi-asumsi tersebut antara lain pengujian kecukupan data dan normal multivariat. Pengujian kecukupan data menggunakan *Kaiser-Meyer-Olkin* (KMO). Sedangkan untuk mendeteksi normal multivariat data menggunakan *Bartlett Test of sphericity* (Bilson, 2005).

KMO diperlukan untuk melihat kecukupan sampel yang dianalisis (*sampling adequacy*). Nilai KMO ini diperoleh dengan membandingkan besarnya koefisien korelasi dengan besarnya koefisien korelasi parsial. Apabila nilai KMO antara 0,5 sampai 1 maka dapat disimpulkan analisis komponen utama tepat digunakan (Bilson, 2005). Bentuk hipotesis KMO, yaitu:

Hipotesis:

H_0 : Ukuran data cukup untuk analisis komponen utama

H_1 : Ukuran data tidak cukup untuk analisis komponen utama

Statistik uji:

$$KMO = \frac{\sum_{i=1}^{p}\sum_{j=1}^{p} r_{ij}^2}{\sum_{i=1}^{p}\sum_{j=1}^{p} r_{ij}^2 + \sum_{i=1}^{p}\sum_{j=1}^{p} a_{ij}^2}\ , \qquad i = 1,2,3\ldots,p\ dan\ j = 1,2,3,\ldots,p \qquad (2.5)$$

dengan r_{ij} adalah koefisien korelasi antara variabel *i* dan *j*, a_{ij} adalah koefisien korelasi parsial antara variabel *i* dan *j*. Kriteria keputusan adalah terima H_0 apabila nilai KMO lebih besar dari 0,5 sehingga dapat disimpulkan ukuran data cukup untuk analisis komponen utama.

Uji *bartlett* bertujuan untuk mengetahui apakah terdapat hubungan antar variabel. Jika variabel $\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_\mathrm{p}$ independen (bersifat saling bebas), maka matriks korelasi antar variabel sama dengan matriks identitas (Olkin, 1995). Bentuk hipotesis uji *bartlett* sebagai berikut:

Hipotesis:

H_0: Matriks korelasi merupakan matriks identitas

H_1: Matriks korelasi bukan merupakan matriks identitas

Statistik uji:

$$\chi^2(\boldsymbol{B}) = -\left[n - \frac{1}{6}(2p+5)\right] \times ln\,|\boldsymbol{R}| \tag{2.6}$$

dengan :

R = matriks korelasi dari masing-masing variabel penjelas

n = jumlah observasi

p = jumlah variabel penjelas

Kriteria keputusan adalah tolak H_0 jika nilai $\chi^2(\boldsymbol{B})$ lebih besar dari $\chi^2_{(\frac{p(p-1)}{2};\alpha)}$.

2.5 Matriks

2.5.1 Matriks Varians-Kovarians dan Matriks Korelasi

Matriks varians-kovarians dari sampel disimbolkan dengan **Σ** atau **S.** vektor mean sampel dari matriks varians-kovarians sampel dapat diperoleh dengan cara sebagai berikut (Anton, 2001):

$$\boldsymbol{\Sigma} = \frac{1}{n-1}\sum_{i=1}^{n}\left(\begin{pmatrix} x_{i1}-\bar{x}_1 \\ x_{i2}-\bar{x}_2 \\ \vdots \\ x_{ip}-\bar{x}_p \end{pmatrix}\begin{pmatrix} x_{i1}-\bar{x}_1 & x_{i2}-\bar{x}_2 & \cdots & x_{ip}-\bar{x}_p \end{pmatrix}\right)$$

$$\boldsymbol{\Sigma} = \frac{1}{n-1}\sum_{i=1}^{n}\begin{pmatrix} (x_{i1}-\bar{x}_1)^2 & (x_{i1}-\bar{x}_1)(x_{i2}-\bar{x}_2) & \cdots & (x_{i1}-\bar{x}_1)(x_{ip}-\bar{x}_p) \\ (x_{i2}-\bar{x}_2)(x_{i1}-\bar{x}_1) & (x_{i2}-\bar{x}_2)^2 & \cdots & (x_{i2}-\bar{x}_2)(x_{ip}-\bar{x}_p) \\ \vdots & \vdots & \ddots & \vdots \\ (x_{ip}-\bar{x}_p)(x_{i1}-\bar{x}_1) & (x_{ip}-\bar{x}_p)(x_{i2}-\bar{x}_2) & \cdots & (x_{ip}-\bar{x}_p)^2 \end{pmatrix}$$

$$\mathbf{\Sigma} = \begin{pmatrix} \frac{1}{n-1}\sum_{i=1}^{n}(x_{i1}-\bar{x}_1)^2 & \frac{1}{n-1}\sum_{i=1}^{n}(x_{i1}-\bar{x}_1)(x_{i2}-\bar{x}_2) & \cdots & \frac{1}{n-1}\sum_{i=1}^{n}(x_{i1}-\bar{x}_1)(x_{ip}-\bar{x}_p) \\ \frac{1}{n-1}\sum_{i=1}^{n}(x_{i2}-\bar{x}_2)(x_{i1}-\bar{x}_1) & \frac{1}{n-1}\sum_{i=1}^{n}(x_{i2}-\bar{x}_2)^2 & \cdots & \frac{1}{n-1}\sum_{i=1}^{n}(x_{i2}-\bar{x}_2)(x_{ip}-\bar{x}_p) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{n-1}\sum_{i=1}^{n}(x_{ip}-\bar{x}_p)(x_{i1}-\bar{x}_1) & \frac{1}{n-1}\sum_{i=1}^{n}(x_{ip}-\bar{x}_p)(x_{i2}-\bar{x}_2) & \cdots & \frac{1}{n-1}\sum_{i=1}^{n}(x_{ip}-\bar{x}_p)^2 \end{pmatrix}$$

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{pmatrix} \tag{2.7}$$

karena $\sigma_{ij} = \sigma_{ji}$ maka $\mathbf{\Sigma}$ dapat dituliskan sebagai berikut:

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{21} & \cdots & \sigma_{p1} \\ \sigma_{12} & \sigma_{22} & \cdots & \sigma_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1p} & \sigma_{2p} & \cdots & \sigma_{pp} \end{pmatrix}$$

Ukuran keeratan hubungan linear antara variabel random A_i dan A_j adalah koefisien korelasi sampel ρ_{ij} yang dinyatakan dalam varians σ_{ii} dan kovarians σ_{ij} pada persamaan berikut ini:

$$\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}}\sqrt{\sigma_{jj}}}$$

Matriks koefisien korelasi adalah matriks simetri berukuran *nxn* yang dapat dinyatakan dalam persamaan 2.8 berikut ini:

$$\rho_{pp} = \begin{pmatrix} \frac{\sigma_{11}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{11}}} & \frac{\sigma_{21}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{11}}} & \cdots & \frac{\sigma_{p1}}{\sqrt{\sigma_{pp}}\sqrt{\sigma_{11}}} \\ \frac{\sigma_{12}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}} & \frac{\sigma_{22}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{22}}} & \cdots & \frac{\sigma_{p2}}{\sqrt{\sigma_{pp}}\sqrt{\sigma_{22}}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\sigma_{1p}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{pp}}} & \frac{\sigma_{2p}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{pp}}} & \cdots & \frac{\sigma_{pp}}{\sqrt{\sigma_{pp}}\sqrt{\sigma_{pp}}} \end{pmatrix}$$

$$= \begin{pmatrix} 1 & \rho_{21} & \cdots & \rho_{p1} \\ \rho_{12} & 1 & \cdots & \rho_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1p} & \rho_{2p} & \cdots & 1 \end{pmatrix} \tag{2.8}$$

2.5.2 Nilai Eigen dan Vektor Eigen

Diberikan sebuah matriks **A**, sebagai berikut:

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$$

Jika **A** adalah matriks berukuran *nxn*, maka vektor tak nol x di dalam R^n dinamakan vektor dari **A**, jika **A**x adalah kelipatan skalar dari x maka **A**$x = \lambda x$. Skalar λ disebut nilai eigen dari **A** dan x disebut vektor eigen yang bersesuaian dengan λ. Untuk menentukan nilai eigen dari matriks **A** yang berukuran *nxn*, dapat dihitung menggunakan persamaan 2.9 berikut ini (Anton, 2001):

$$\det(\mathbf{A} - \lambda\mathbf{I}) = 0 \tag{2.9}$$

Persamaan 2.9 disebut dengan persamaan karakteristik.

Untuk setiap vektor eigen x dari **A**, $e = \frac{x}{||x||}$ merupakan vektor eigen dari **A** yang bersesuaian dengan λ yang sama dan $|| x || = 1$. Proses seperti ini dinamakan proses normalisasi dan vektor yang didapatkan adalah vektor yang telah dinormalisasikan.

2.6 Analisis Komponen Utama

Analisis Komponen Utama (AKU) merupakan suatu teknik statistika untuk mentransformasi variabel-variabel asli yang saling berkorelasi satu dengan yang lain menjadi satu set variabel baru yang tidak berkorelasi lagi. Jadi analisis komponen utama berguna untuk mereduksi data**,** sehingga lebih mudah untuk menginterpretasikan data-data tersebut. Menurut Johnson (2002) AKU terkonsentrasi pada penjelasan struktur variansi dan kovariansi melalui suatu kombinasi linear variabel-variabel asal, dengan tujuan utama melakukan reduksi data dan membuat interpretasi. Analisis komponen utama lebih baik digunakan jika variabel-variabel asal saling berkorelasi

Jika vektor random **X**' = [X_1, X_2, …, X_p] mempunyai matriks varian-kovarian $\sum$ dengan nilai eigen $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$, maka kombinasi linear didefinisikan pada persamaan (2.10) berikut ini (Johnson,2002):

$$
\begin{aligned}
\mathrm{w}_1 &= \boldsymbol{e}_1'\boldsymbol{X} = e_{11}X_1 + e_{12}X_2 + \cdots + e_{1p}X_p \\
\mathrm{w}_2 &= \boldsymbol{e}_2'\boldsymbol{X} = e_{21}X_1 + e_{22}X_2 + \cdots + e_{2p}X_p \\
&\vdots \\
\mathrm{w}_p &= \boldsymbol{e}_p'\boldsymbol{X} = e_{p1}X_1 + e_{p2}X_2 + \cdots + e_{pp}X_p
\end{aligned}
\tag{2.10}
$$

$$
\boldsymbol{e}_j'\boldsymbol{X} = \begin{pmatrix} e_{11} & e_{12} & \cdots & e_{1p} \\ e_{21} & e_{22} & \cdots & e_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ e_{p1} & e_{p2} & \cdots & e_{pp} \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix}
$$

dengan w_1 adalah komponen pertama yang memenuhi maksimum nilai $\boldsymbol{e}_1' \sum \boldsymbol{e_1} = \lambda_1$, w_2 adalah komponen kedua yang memenuhi sisa keragaman selain komponen pertama dengan memaksimumkan nilai $\boldsymbol{e}_2' \sum \boldsymbol{e_2} = \lambda_2$, dan w_p adalah komponen ke-p yang memenuhi sisa keragaman selain dari $\mathrm{w}_1, \mathrm{w}_2, \ldots, \mathrm{w}_{p-1}$ dengan memaksimumkan nilai $\boldsymbol{e}_p' \sum \boldsymbol{e_p} = \lambda_p$. Syarat untuk membentuk komponen utama yang merupakan kombinasi linear dari variabel *X* agar mempunyai variansi maksimum adalah dengan memilih vektor eigen yaitu $\boldsymbol{e} = (\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_p)$ sedemikian hingga Var (w_i) = $\boldsymbol{e}_i' \sum \boldsymbol{e_i}$ maksimum dan $\boldsymbol{e_i}'\boldsymbol{e_i} = 1$ (Hair, 1998).

Penurunan komponen utama dari matriks korelasi dilakukan apabila data sudah terlebih dahulu ditransformasikan ke dalam bentuk baku Z. Transformasi ini dilakukan terhadap data yang satuan pengamatannya tidak sama. Bila variabel yang diamati ukurannya pada skala dengan perbedaan yang sangat lebar atau satuan ukurannya tidak sama, maka variabel tersebut perlu dibakukan (*standardized*). Sebaliknya matriks varians-kovarians dilakukan terhadap data yang satuan pengamatannya sama (Hardle, 2003)

Kriteria pemilihan komponen utama menggunakan persentase variansi kumulatif terhadap total variansi. Persentase variansi kumulatif yang dijelaskan komponen utama ke-*i* dapat dihitung menggunakan persamaan (2.11) berikut ini (Johnson, 2002):

$$
\frac{\lambda_i}{\sum_{i=1}^{p}\lambda_i} \times 100\% = \frac{\lambda_i}{\lambda_1+\lambda_2+\cdots+\lambda_p} \times 100\% \text{, untuk } i = 1, 2, \ldots, p \tag{2.11}
$$

Pemilihan komponen utama pada dasarnya sesuai kebutuhan peneliti karena tidak ada aturan khusus dalam pemilihan jumlah komponen utama. Menurut Bilson (2005) jumlah komponen utama dapat ditentukan dengan melihat persentase variansi kumulatif yang dipilih mampu menerangkan total variansi data sekitar 70% sampai 85%. Selain itu, jumlah komponen utama juga dapat diketahui dengan melihat pola penurunan nilai eigen pada *scree plot*. Penurunan yang tajam pada *scree plot* menunjukkan perubahan nilai eigen yang besar. *Scree plot* adalah plot antara nilai eigen dengan *i*. Cara lain dalam pemilihan komponen utama juga dapat ditentukan berdasarkan nilai eigen yang lebih dari satu. Hal tersebut dikarenakan nilai eigen yang mendekati nol dianggap tidak memberikan pengaruh yang penting.

2.7 Penentuan Bandwidth pada Regresi Terboboti Geografis

Definisi *bandwidth* secara teoritis adalah lingkaran dengan radius *h* dari titik lokasi tertentu yang digunakan sebagai dasar untuk menentukan bobot setiap pengamatan terhadap model regresi pada lokasi tersebut. Pengamatan yang terletak dekat dengan lokasi *i* akan lebih berpengaruh dalam pendugaan parameter model di setiap lokasi. Pengamatan yang berada dalam radius *h* dianggap berpengaruh dalam membentuk parameter di titik lokasi pengamatan ke-*i*. Pemilihan *bandwidth* menjadi sangat penting karena akan mempengaruhi keakuratan model data, yang mengatur variasi dan bias model (Maulani,2013).

Nilai *bandwidth* yang sangat kecil akan mengakibatkan penaksiran parameter pada lokasi pengamatan ke-i semakin bergantung pada titik lokasi pengamatan lain yang memiliki jarak terdekat dengan lokasi pengamatan ke-i, sehingga varians yang dihasilkan semakin besar. Sebaliknya, jika nilai bandwidth sangat besar maka akan mengakibatkan bias yang semakin besar, sehingga model yang diperoleh terlalu halus (Maulani,2013).

Salah satu metode untuk menentukan *bandwidth* optimum adalah validasi silang atau *cross validation* (CV). *Bandwidth* optimum adalah *bandwidth* yang menghasilkan nilai CV minimum. Apabila nilai-nilai penduga y_i merupakan fungsi dari *bandwidth* (h) ditulis $\hat{y}_i(h)$, maka nilai *bandwidth* dengan metode CV diperoleh dengan menghilangkan observasi ke-i dalam menduga nilai y_i pada model. Secara matematis CV dapat dituliskan sebagai berikut (Fotheringham, 2002):

$$CV(h) = \sum_{i=1}^{n} \left(y_i - \hat{y}_{\neq i}(h)\right)^2 \tag{2.12}$$

dengan, $\hat{y}_{\neq i}(h)$ adalah penduga y_i dengan pengamatan di lokasi (u_i, v_i) dihilangkan dari proses pendugaan, y_i merupakan pengamatan ke-i, dan n merupakan jumlah sampel.

Proses untuk mendapatkan *bandwidth* yang meminimumkan nilai CV bisa dilakukan dengan menggunakan teknik *Golden Section Search* (Fotheringham, 2002). Proses ini dilakukan dengan cara mengevaluasi fungsi dengan tiga nilai yang berbeda, misalnya a, b, dan c, dengan a < b < c, a merupakan batas bawah nilai *bandwidth* yang mungkin dan c merupakan batas atas nilai *bandwidth* yang mungkin. Nilai a diperoleh dari nilai minimum d_{ij} sedangkan c diperoleh dari nilai maksimum d_{ij}. Nilai fungsi yang dihasilkan pada tiga titik tersebut adalah f(a), f(b), dan f(c), yang disebut juga sebagai triplet. Fungsi tersebut dievaluasi lagi pada suatu nilai baru d yang bisa ditentukan di antara a dan b atau di antara b dan c sehingga menghasilkan nilai fungsi baru, yaitu f(d). Kemudian buang salah satu dari nilai a atau c untuk membentuk triplet baru. Aturan yang digunakan pada teknik *Golden Section Search* adalah:

Jika f(b) < f(d) : triplet baru yang digunakan adalah a < b < d
Jika f(b) > f(d) : triplet baru yang digunakan adalah b < d < c

Proses tersebut berulang sampai dengan dua nilai f(d) yang dihasilkan mendekati sama atau selisihnya lebih kecil daripada suatu nilai yang ditentukan, misal $1\text{x}10^{-6}$, atau sampai suatu nilai iterasi maksimum yang diperbolehkan.

2.8 Regresi Terboboti Geografis

Model regresi terboboti geografis (RTG) pertama kali diperkenalkan oleh Fotheringham pada tahun 1967. Model RTG merupakan pengembangan dari model regresi global yang ide dasarnya diambil dari regresi non parametrik. Variabel prediktor pada RTG masing-masing koefisien regresinya bergantung pada lokasi yang mana data tersebut diamati. RTG merupakan suatu metode regresi yang menghasilkan penduga parameter yang hanya dapat digunakan untuk memprediksi setiap titik atau lokasi data tersebut diamati dan dikumpulkan (Fotheringham, 2002).

Fotheringham (2002) dan Huang (2010) menjelaskan bentuk umum dari model RTG secara matematis seperti yang ditunjukkan pada persamaan 2.13 berikut ini:

$$y_i = \beta_0(u_i, v_i) + \sum_{j=1}^{p} \beta_j\,(u_i, v_i)x_{ij} + \varepsilon_i \quad ; i = 1,2, \dots, n \tag{2.13}$$

dengan :

y_i = nilai observasi variabel respon lokasi ke-*i*

x_{ij} = nilai observasi variabel prediktor *j* pada lokasi ke-*i*

$\beta_0(u_i, v_i)$ = nilai intersep model RTG

$\beta_j(u_i, v_i)$ = parameter regresi untuk setiap lokasi ke-*i*

(u_i, v_i) = titik koordinat (lintang, bujur) pada lokasi ke-*i*

ε_i = galat pada lokasi ke-*i*

Setiap parameter yang ada pada model RTG dihitung berdasarkan titik setiap lokasi geografis sehingga nilai parameternya memiliki perbedaan antara satu wilayah dengan wilayah lainya. Pendugaan parameter RTG dapat dilakukan dengan metode *Weighted Least Square* (WLS) dengan memberikan pembobot yang berbeda untuk masing-masing lokasi data (Fotheringham, 2002).

Misalkan pembobot untuk setiap titik lokasi pengamatan (u_i, v_i) adalah W_{ij} dengan $j = 1,2, \dots, n$, maka estimasi parameter model dilakukan dengan meminimumkan jumlah kuadrat *error*-nya, pada persamaan 2.13 yang telah diberi pembobot W_{ij} , sehingga diperoleh:

$$\sum_{j=1}^{n} W_j(u_i, v_i)\,\varepsilon_i^2 = \sum_{j=1}^{n} W_j(u_i, v_i)\Big(y_i - \beta_0(u_i, v_i) - \sum_{j=1}^{p} \beta_j\,(u_i, v_i)x_{ij}\Big)^2 \tag{2.14}$$

Fungsi RTG dalam bentuk matriks,

$$\boldsymbol{Y} = \boldsymbol{X\beta}(u_i, v_i) + \boldsymbol{\varepsilon}$$

$$\boldsymbol{W}(u_i, v_i) = diag[(W_1(u_i, v_i), W_2(u_i, v_i), \dots \ , W_n(u_i, v_i)]$$

dengan :

$$\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)'$$

Solusi dari Persamaan 2.14 dalam bentuk matriks adalah,

$$\begin{aligned}\boldsymbol{\varepsilon}'\boldsymbol{W}(u_i,v_i)\boldsymbol{\varepsilon} &= [\boldsymbol{Y} - \boldsymbol{X\beta}(u_i,v_i)]'\boldsymbol{W}(u_i,v_i)[\,\boldsymbol{Y} - \boldsymbol{X\beta}(u_i,v_i)]\\ &= \boldsymbol{Y}'\boldsymbol{W}(u_i,v_i)\boldsymbol{Y} - \boldsymbol{Y}'\boldsymbol{W}(u_i,v_i)\,\boldsymbol{X\beta}(u_i,v_i) -\\ &\quad \boldsymbol{\beta}'(u_i,v_i)\boldsymbol{X}'\boldsymbol{W}(u_i,v_i)\boldsymbol{Y} + \boldsymbol{\beta}'(u_i,v_i)\boldsymbol{X}'\boldsymbol{W}(u_i,v_i)\boldsymbol{X\beta}(u_i,v_i)\end{aligned}$$

karena $\boldsymbol{X\beta}(u_i,v_i) = \boldsymbol{\beta}'(u_i,v_i)\boldsymbol{X}'$ maka menjadi :

$$\begin{aligned}\boldsymbol{\varepsilon}'\boldsymbol{W}(u_i,v_i)\,\boldsymbol{\varepsilon} &= \boldsymbol{Y}'\boldsymbol{W}(u_i,v_i)\boldsymbol{Y} - 2\boldsymbol{\beta}'(u_i,v_i)\boldsymbol{X}'\boldsymbol{W}(u_i,v_i)\boldsymbol{Y}\\ &\quad +\boldsymbol{\beta}'(u_i,v_i)\boldsymbol{X}'\boldsymbol{W}(u_i,v_i)\boldsymbol{X\beta}(u_i,v_i) \qquad (2.15)\end{aligned}$$

Persamaan 2.15 diturunkan terhadap matriks $\boldsymbol{\beta}'(u_i,v_i)$ dan hasilnya disamadengankan dengan nol, sehingga:

$$\left.\frac{\partial \boldsymbol{\varepsilon}'\boldsymbol{W}(u_i,v_i)\,\boldsymbol{\varepsilon}}{\partial \boldsymbol{\beta}'(u_i,v_i)}\right|_{\boldsymbol{\beta}(u_i,v_i)=\widehat{\boldsymbol{\beta}}(u_i,v_i)} = 0$$

$$-2\boldsymbol{X}'\boldsymbol{W}(u_i,v_i)\boldsymbol{Y} + 2\boldsymbol{X}'\boldsymbol{W}(u_i,v_i)\boldsymbol{X\beta}(u_i,v_i) = 0$$

$$\boldsymbol{X}'\boldsymbol{W}(u_i,v_i)\boldsymbol{X\beta}(u_i,v_i) = \boldsymbol{X}'\boldsymbol{W}(u_i,v_i)\boldsymbol{Y}$$

$$(\boldsymbol{X}'\boldsymbol{W}(u_i,v_i)\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W}(u_i,v_i)\boldsymbol{X\beta}(u_i,v_i) = (\boldsymbol{X}'\boldsymbol{W}(u_i,v_i)\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W}(u_i,v_i)\boldsymbol{Y}$$

$$\widehat{\boldsymbol{\beta}}(u_i,v_i) = (\boldsymbol{X}'\boldsymbol{W}(u_i,v_i)\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W}(u_i,v_i)\boldsymbol{Y}$$

Estimasi parameter dari model RTG untuk setiap lokasi adalah:

$$\widehat{\boldsymbol{\beta}}(u_i,v_i) = (\boldsymbol{X}'\boldsymbol{W}(u_i,v_i)\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W}(u_i,v_i)\boldsymbol{Y} \qquad (2.16)$$

Misalkan $\boldsymbol{x}_{\boldsymbol{i}}' = (1, x_{i1}, x_{i2}, \dots, x_{ip})$ adalah elemen baris ke-*i* pada matriks **X,** dan $\widehat{\boldsymbol{\beta}}(u_i,v_i)$ adalah vektor estimasi parameter pada lokasi (u_i,v_i) maka diperoleh estimasi model RTG untuk pengamatan ke-*i* sebagai berikut:

$$\begin{aligned}\widehat{\boldsymbol{y}}_{\boldsymbol{i}} &= \boldsymbol{x}_{\boldsymbol{i}}' \cdot \widehat{\boldsymbol{\beta}}(\boldsymbol{u_i},\boldsymbol{v_i})\\ &= \boldsymbol{x}_{\boldsymbol{i}}' \cdot (\boldsymbol{X}'\boldsymbol{W}(u_i,v_i)\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W}(u_i,v_i)\boldsymbol{Y}\end{aligned}$$

2.9 Pemilihan Pembobot

Pembobot pada model RTG memiliki peran yang sangat penting karena nilai pembobot mewakili lokasi data yang diamati satu dengan yang lain. Penaksiran parameter titik spasial (u_i, v_i) mempunyai pengaruh lebih besar apabila titik-titiknya saling berdekatan pada lokasi *i*. Oleh sebab itu, pemilihan pembobot spasial sangat penting dalam analisis spasial karena nilai pembobot mewakili letak data yang satu dengan data yang lain sehingga ketepatan cara pembobotan dibutuhkan.

Matriks pembobot pada RTG merupakan matriks pembobot berbasis pada kedekatan titik lokasi pengamatan ke-i dengan titik lokasi pengamatan lainnya. Pengamatan terdekat ke titik lokasi pengamatan ke-i umumnya diasumsikan memiliki pengaruh paling besar terhadap penaksiran parameter di titik lokasi pengamatan ke-*i*. Oleh karena itu, matriks pembobot $\boldsymbol{W}(u_i, v_i)$ akan semakin besar seperti jarak yang semakin dekat (Maulani,2013).

Untuk mendapatkan matriks pembobot pada lokasi *i* terletak pada koordinat (u_i, v_i) yaitu lintang dan bujur. Perlu terlebih dahulu menentukan fungsi pembobot yang akan digunakan. Jika lokasi *j* terletak pada titik koordinat (u_j, v_j) maka akan diperoleh jarak *euclid* antara lokasi *i* dan lokasi *j* dengan menggunakan persamaan 2.17 berikut (Fotheringham, 2014):

$$d_{ij} = \sqrt{\left(u_i - u_j\right)^2 - \left(v_i - v_j\right)^2} \tag{2.17}$$

Misalkan matriks pembobot untuk setiap titik lokasi pengamatan ke-i adalah $\boldsymbol{W}_{ij}$ dan dinyatakan sebagai berikut:

$$\boldsymbol{W}_{ij} = \begin{pmatrix} W_{i1} & 0 & \dots & 0 \\ 0 & W_{i2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & W_{in} \end{pmatrix}$$

Terdapat beberapa jenis fungsi pembobot dalam RTG yang dapat digunakan untuk menentukan matriks pembobot (Wheeler, 2010), diantaranya:

1. Fungsi Invers Jarak

   Fungsi tersebut dinotasikan sebagai berikut:

   $$W_{ij} = \begin{cases} 1 & jika\, d_{ij} < h \\ 0 & jika\, d_{ij} \geq h \end{cases}$$

Fungsi invers jarak akan memberi bobot nol ketika lokasi *j* berada diluar radius *h* dari lokasi *i*, sedangkan apabila lokasi *j* berada dalam radius *h* maka akan mendapat bobot satu.

2. Fungsi Kernel Gaussian

   Bentuk fungsi kernel gaussian adalah

   $$W_{ij} = exp\left(-\frac{1}{2}(d_{ij}/h)^2\right)$$

   Fungsi kernel gaussian akan memberi bobot yang akan semakin menurun mengikuti fungsi atau kurva gaussian ketika d_{ij} semakin besar.

3. Fungsi Kernel *Bisquare*

   Fungsi tersebut dapat dinotasikan sebagai berikut:

   $$W_{ij} = \begin{cases} \left(1-\left(\frac{d_{ij}}{h}\right)^2\right)^2 & jika\ d_{ij} < h \\ 0 & jika\ d_{ij} \geq h \end{cases}$$

   Fungsi kernel *bisquare* akan memberi bobot nol jika lokasi *j* berada diluar radius *h* dari lokasi *i*, sedangkan apabila lokasi *j* berada di dalam radius *h* maka akan diberi bobot yang mengikuti fungsi kernel *bisquare*.

4. Fungsi Kernel Adaptif Gaussian

   Fungsi tersebut dapat dinotasikan sebagai berikut:

   $$W_{ij} = exp\left(-\frac{1}{2}(d_{ij}/h_i)^2\right) \tag{2.18}$$

5. Fungsi Kernel Adaptif *Bisquare*

   Fungsi tersebut dapat dinotasikan sebagai berikut:

   $$W_{ij} = \begin{cases} \left(1-\left(\frac{d_{ij}}{h_i}\right)^2\right)^2 & jika\ d_{ij} < h \\ 0 & jika\ d_{ij} \geq h \end{cases}$$

Pada penelitian ini fungsi yang digunakan dalam menentukan matriks pembobot adalah fungsi kernel adaptif gaussian. Hal ini disebabkan kemampuan fungsi kernel adaptif gaussian yang dapat disesuaikan dengan kondisi titik-titik pengamatan. Jika titik-titik lokasi pengamatan tersebar secara padat di sekitar lokasi pengamatan ke-*i* maka *bandwidth* yang diperoleh relatif sempit. Sebaliknya jika titik-titik lokasi pengamatan memiliki jarak yang relatif jauh dari titik lokasi pengamatan ke-*i* maka *bandwidth* yang diperoleh akan semakin luas (Wheeler, 2010).

2.10 Regresi Komponen Utama Terboboti Geografis

Regresi Komponen Utama Terboboti Geografis (RKUTG) merupakan gabungan dari metode RTG dan KU. Setelah dilakukan uji heterogenitas spasial dan diketahui terdapat perbedaan karakteristik suatu wilayah dengan wilayah lainnya maka metode RTG dapat digunakan. Perbedaan karakteristik suatu wilayah meliputi adanya perbedaan sumberdaya, lingkungan dan geografi. Apabila terdapat multikolinearitas lokal pada model RTG, selanjutnya dilakukan analisis komponen utama untuk mereduksi variabel-variabel bebas yang saling berkorelasi dengan menggunakan metode KU.

2.10.1 Uji Kesesuaian Model RKUTG

Uji kesesuaian model dilakukan untuk mengetahui apakah faktor geografis berpengaruh signifikan atau tidak terhadap model.

Bentuk hipotesis pengujian:

H_0 : $\beta_j(u_i, v_i) = \beta_j$; $j = 1, 2, \dots, p,\ i = 1,2, \dots, n$

(Tidak ada perbedaan yang signifikan antara model RTG dan RKUTG)

H_1 : minimal ada satu $\beta_j(u_i, v_i) \neq \beta_j$; $j = 1,2, \dots, p,\ i = 1,2, \dots, n$

(Terdapat perbedaan yang signifikan antara model RTG dan RKUTG)

Taraf kepercayaan yang digunakan dalam penelitian ini adalah 5%, hal ini menunjukkan bahwa hasil penelitian dapat dipercaya kebenarannya sebesar 95% dan memiliki kesalahan paling besar 5%.

Statistik uji yang digunakan adalah:

$$F_{hitung} = \frac{SSE(H_0)/df_1}{SSE(H_1)/df_2} \quad (2.19)$$

dengan:

$SSE(H_0) = \boldsymbol{Y}'(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y}$

$SSE(H_1) = \boldsymbol{Y}'(\boldsymbol{I} - \boldsymbol{S})'(\boldsymbol{I} - \boldsymbol{S})\boldsymbol{Y}$

$\boldsymbol{H} \quad = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$

$df_1 \quad = n - p - 1$

$df_2 \quad = \left(n - 2tr(\boldsymbol{S}) + tr(\boldsymbol{S}'\boldsymbol{S})\right)$

$$\boldsymbol{S} = \begin{pmatrix} x_1'[\boldsymbol{X}'\boldsymbol{W}(u_1,v_1)\boldsymbol{X}]^{-1}\boldsymbol{X}'\boldsymbol{W}(u_1,v_1) \\ x_2'[\boldsymbol{X}'\boldsymbol{W}(u_2,v_2)\boldsymbol{X}]^{-1}\boldsymbol{X}'\boldsymbol{W}(u_2,v_2) \\ \vdots \\ x_n'[\boldsymbol{X}'\boldsymbol{W}(u_n,v_n)\boldsymbol{X}]^{-1}\boldsymbol{X}'\boldsymbol{W}(u_n,v_n) \end{pmatrix}$$

I = matriks identitas

Kriteria uji:

Tolak H_0 jika $F_{hitung} > F_{tabel}\left(F_{(\alpha;\, df_1, df_2)}\right)$
atau dengan kata lain model RKUTG merupakan model yang lebih baik daripada model RTG.

2.10.2 Uji Parsial Parameter Model RKUTG

Uji parsial parameter RKUTG dilakukan untuk mengetahui variabel-variabel penjelas yang berpengaruh signifikan terhadap variabel respon. Bentuk hipotesis pengujian:

H_0 : $\beta_j(u_i,v_i) = 0$; j=1, 2, ... , p dan i=1,2,…,n (variabel x_j pada lokasi ke-i tidak mempengaruhi variabel respon di lokasi tersebut)

H_1 : $\beta_j(u_i,v_i) \neq 0$; j=1, 2, ... ,p dan i=1,2,…,n (variabel x_j pada lokasi ke-i mempengaruhi variabel respon di lokasi tersebut)

Statistik uji yang digunakan:

$$t = \frac{\hat{\beta}_j(u_i,v_i)}{\hat{\sigma}\sqrt{g_{jj}}} \tag{2.20}$$

dengan g_{jj} adalah elemen diagonal ke-j dari matriks $\boldsymbol{LL}'$ dan $\hat{\sigma} = \sqrt{\frac{SSE(H_0)}{\delta_1}}$,

$\boldsymbol{L} = (\boldsymbol{X}'\boldsymbol{W}(u_i,v_i)\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W}(u_i,v_i)$

$\delta_1 = tr(\boldsymbol{S}'\boldsymbol{S})$.

Kriteria Uji:

$$|t| = \begin{cases} \geq \ t_{\left(\frac{\alpha}{2}, df_2\right)}, & \text{tolak } H_0 \\ < t_{\left(\frac{\alpha}{2}, df_2\right)}, & \text{terima } H_0 \end{cases}$$

Jika H_0 ditolak maka dapat disimpulkan bahwa parameter $\beta_j(u_i, v_i)$ berpengaruh terhadap model RKUTG.

2.11 Pemilihan Model Terbaik

Metode yang digunakan untuk mendapatkan model terbaik adalah *Akaike Information Criterion* (AIC), secara matematis dituliskan sebagai berikut (Fotheringham, 2002):

$$AIC = 2\, n \ln(\hat{\sigma}) + n \ln(2\pi) + n + tr(\boldsymbol{S}) \qquad (2.21)$$

dengan $\hat{\sigma}$ merupakan nilai estimator standar deviasi dari bentuk residual dan **S** merupakan matriks proyeksi dari model.

2.12 Produk Domestik Regional Bruto

Perhitungan PDRB telah menjadi bagian yang sangat penting dalam makro ekonomi, khususnya tentang analisis perekonomian suatu daerah. Hasil perhitungan PDRB ini memberikan kerangka dasar yang digunakan untuk mengukur aktivitas ekonomi yang terjadi dan berlangsung dalam suatu kegiatan perekonomian. Angka-angka PDRB tersebut sebagai indikator ekonomi makro dan juga sebagai landasan evaluasi kinerja perekonomian, dan penyusutan berbagai kebijakan. Indikator ekonomi ini juga memberikan gambaran aliran seluruh faktor-faktor produksi yang digunakan oleh perekonomian untuk menghasilkan nilai tambah barang dan jasa.

PDRB merupakan jumlah nilai tambah yang dihasilkan untuk seluruh daerah usaha dan jasa dalam suatu daerah, menerapkan jumlah seluruh nilai barang dan jasa akhir yang dihasilkan seluruh unit ekonomi (BPS, 2019). PDRB dapat didefinisikan sebagai jumlah nilai tambah yang dihasilkan oleh seluruh unit usaha di suatu daerah. Penyusutan PDRB dapat dilakukan melalui 3 (tiga) pendekatan yaitu pendekatan produksi, pengeluaran, dan pendapatan yang disajikan atas dasar harga berlaku.

PDRB harga berlaku bertujuan untuk melihat struktur perekonomian yang menunjukkan kemampuan sumber daya ekonomi yang dihasilkan oleh suatu daerah. Selain itu, kondisi perekonomian secara keseluruhan di setiap daerah juga dapat dilihat dari seberapa besar jumlah belanja daerah pada daerah bersangkutan (BPS, 2019). Belanja daerah merupakan bentuk rangsangan yang dilakukan oleh pemerintah terhadap perekonomian daerah. Jika nilai belanja daerah yang dialokasikan untuk pembangunan besar, maka akan meningkatkan kesejahteraan penduduk. Hal ini berarti kondisi ekonomi di daerah tersebut akan meningkat.

Struktur perekonomian dalam suatu daerah dapat dipengaruhi oleh beberapa faktor yang dapat menyebabkan terjadinya kenaikan pertumbuhan PDRB. Berikut kriteria yang menjadikan sebagai indikator dalam pertumbuhan PDRB menurut standar BPS yaitu indeks pembangunan manusia, angka harapan hidup, rata-rata lama sekolah, pengeluaran perkapita, penduduk miskin, jumlah pertokoan dan pasar, rumah tangga menggunakan listrik, angka melek huruf, jumlah penginapan dan hotel, rumah tangga menggunakan gas, dan tingkat Pendidikan.

Penelitian ini bertujuan untuk menerapkan metode model analisis komponen utama regresi terboboti secara geografis pada data PDRB yang mengalami multikolinearitas. Menurut standar BPS, tingkat PDRB dapat dipengaruhi oleh 11 faktor, dari 11 faktor tersebut asumsi multikolinearitas tidak terpenuhi. Oleh karena itu, dilakukan pengurangan faktor dan kombinasi untuk memperoleh syarat data yang multikolinearitas. Sehingga diperoleh 7 faktor yang memenuhi syarat data multikolinearitas.

BAB III
METODE PENELITIAN

3.1 Sumber Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari data dan informasi produk domestik regional bruto (PDRB) kabupaten/kota Sulawesi Selatan tahun 2019. Data diambil dari publikasi BPS pada situs *https://sulsel.bps.go.id/publication.* Data penelitian ini menggunakan data kuantitatif dengan satuan persentase yang diolah menggunakan *software* R, dan menggunakan data letak astronomi yang meliputi letak lintang dan letak bujur tiap kabupaten/kota di Sulawesi Selatan sebagai faktor pembobot geografis. Lokasi penelitian terdiri dari 24 kabupaten/kota di Provinsi Sulawesi Selatan.

3.2 Deskripsi Variabel

Variabel yang digunakan dalam penelitian ini adalah:

a. Variabel Respon (Y)

Y: Persentase Produk Domestik Regional Bruto (PDRB) kabupaten/kota di Provinsi Sulawesi Selatan tahun 2019.

b. Variabel Prediktor (X)

X_1: Persentase Indeks Pembangunan Manusia (IPM) kabupaten/kota di Provinsi Sulawesi Selatan tahun 2019.

X_2: Persentase angka harapan hidup Kabupaten/Kota di Provinsi Sulawesi Selatan tahun 2019.

X_3: Persentase rata-rata lama sekolah Kabupaten/Kota di Provinsi Sulawesi Selatan tahun 2019.

X_4: Persentase pengeluaran perkapita Kabupaten/Kota di Provinsi Sulawesi Selatan tahun 2019.

X_5: Persentase penduduk miskin Kabupaten/Kota di Provinsi Sulawesi Selatan tahun 2019.

X_6: Persentase rumah tangga menggunakan listrik Kabupaten/Kota di Provinsi Sulawesi Selatan tahun 2019.