PERFORMA MODEL *STATISTICAL DOWNSCALING* DENGAN PEUBAH *DUMMY* BERDASARKAN *K-MEANS* DAN *AVERAGE LINKAGE*

SKRIPSI

FITRI ANNISA

H121 13 002

PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

MAKASSAR

MEI 2020

PERFORMA MODEL *STATISTICAL DOWNSCALING* DENGAN PEUBAH *DUMMY* BERDASARKAN *K-MEANS* DAN *AVERAGE LINKAGE*

SKRIPSI

Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Sains pada Program Studi Statistika Departemen Statistika

Fakultas Matematika dan Ilmu Pengetahuan Alam

Universitas Hasanuddin Makassar

FITRI ANNISA

H 121 13 002

PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUANALAM

UNIVERSITAS HASANUDDIN

MAKASSAR

MEI 2020

LEMBAR PERNYATAAN KEOTENTIKAN

Saya yang bertandatangan di bawah ini menyatakan dengan sungguh-sungguh bahwa skripsi yang saya buat dengan judul

PERFORMA MODEL *STATISTICAL DOWNSCALING*

DENGAN PEUBAH *DUMMY* BERDASARKAN

K-MEANS* DAN *AVERAGE LINKAGE

Adalah benar hasil karya saya sendiri, bukan hasil plagiat dan belum pernah dipublikasikan dalam bentuk apapun.

Makassar, 14 Mei 2020

METERAI TEMPEL

D7AHF556713386

6000

ENAM RIBU RUPIAH

Fitri Annisa

NIM. H121 13 002

PERFORMA MODEL *STATISTICAL DOWNSCALING*

DENGAN PEUBAH *DUMMY* BERDASARKAN

K-MEANS* DAN *AVERAGE LINKAGE

Disetujui Oleh :

Pembimbing Utama	Pembimbing Pendamping
Drs. Raupong, M.Si	**Sitti Sahriman, S.Si., M.Si**
NIP. 19621015 198810 1 001	**NIP. 19881018 201504 2 002**

Ketua Program Studi

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

NIP. 19720117 199703 2 002

Pada Tanggal : 14 Mei 2020

HALAMAN PENGESAHAN

Skripsi ini diajukan oleh :

Nama : Fitri Annisa

NIM : H 121 13 002

Program Studi : Statistika

Judul Skripsi : Performa Model *Statistical Downscaling* dengan Peubah *Dummy* Berdasarkan *K-means* dan *Average Linkage*

Telah berhasil dipertahankan dihadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Sains pada Program Studi Statistika Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

DEWAN PENGUJI

Tanda Tangan

1. **Ketua : Drs. Raupong, M.Si.** ( )

2. **Sekertaris : Sitti Sahriman,S.Si., M.Si.** ( )

3. **Anggota : Andi Kresna Jaya,S.Si.,M.Si.** ( )

4. **Anggota : Dr. Dr Georgina Maria Tinungki, M.Si.** ( )

Ditetapkan di : Makassar

Tanggal : 14 Mei 2020

KATA PENGANTAR

Segala puji bagi Allah *Subhanahu Wa ta'alaRabb* semesta alam, shalawat serta salam semoga senantiasa tercurahkan kepada Nabi yang paling dimuliakan, pemimpin orang-orang bertakwa yakni Rasulullah Muhammad *Shallallahu AlaihiWasallam* dan kepada para keluarga serta shahabat beliau yang senantiasa kitarindukan perjumpaan dengannya. Amma ba'du.

"*Maka sesungguhnya bersama kesulitan ada kemudahan.Sesungguhnya bersama kesulitan ada kemudahan*"(QS. Al-Insyirah : 5-6)

Alhamdulillah, semua kemudahan yang penulis dapatkan tidak lepas dari pertolongan Allahdan doa dari orang-orang yang tulus dari penulisan skripsi ini, akhirnya skripsi dengan judul **"Performa Model *Statistical Downscaling* dengan Peubah *Dummy* Berdasarkan *K-means dan Average Linkage*"** yang disusun sebagai salah satu syarat akademik untuk meraih gelar sarjana pada Program Studi Statistika Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin ini dapat dirampungkan. Dalam penulisan skripsi ini, penulis dengan segala keterbatasan kemampuan dan pengetahuan dapat melewati segala hambatan dan masalah berkat bantuan dan dorongan dari berbagai pihak. Penulis berharap skripsi ini bisa memberikan tambahan pengetahuan bagi pembelajar statistika.

Tidak lupa penulis mengucapkan terimakasih kepada seluruh pihak yang senantiasa membantu baik berupa materi, tenaga dan dukungan moral selama proses penyelesaian tulisan ini :

1. **Rektor Universitas Hasanuddin**, **Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam**, **Ketua Departemen Statistika**, **segenap dosen pengajar** dan **staf Jurusan Statistika** serta **staf Fakultas MIPA** yang telah membekali ilmu dan kemudahan-kemudahan kepada penulis dalam berbagai hal selama menjadi mahasiswa di Jurusan Matematika.
2. Bapak **Drs.Raupong, M.Si** selaku dosen pembimbing utama yang telah dengan sabar dan ikhlas meluangkan begitu banyak waktunya untuk membimbing dan memberikan masukan serta motivasi dalam penulisan skripsi ini.

3. Ibu **Sitti Sahriman,S.Si., M.Si** selaku dosen pembimbing pendamping yang juga telah membekali ilmu selama bimbingan dan memberikan masukan serta motivasi dalam penulisan skripsi ini.
4. Bapak **Andi Kresna Jaya, S.Si., M.Si.** dan **Dr. Dr. Georgina Maria Tinungki, M.Si**, selaku penguji yang telah memberikan kritik dan saran yang membangun untuk penulisan skripsi ini. Serta yang telah banyak memberikanmasukan yang sangat berharga dalam perbaikan skripsi ini. Terima kasih jugaatas waktu yang telah diluangkan untuk memberikan banyak nasehat selama penulis menjalani perkuliahan di Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

Ucapan terimakasih juga penulis sampaikan kepada orang-orang yang telah berperan besar serta istimewa kepada :

1. Kedua orang tua tercinta**, Rury Maryun** dan **Linda**, sebagai madrasah pertama yang telah banyak memberikan pelajaran serta pendidikan sebagai bekal kehidupan. Mereka telah memberikan kasih sayang, doa, dan motivasi yang besar. Semoga ini bisa menjadi salah satu amal jariyah bagi mereka.
2. Saudara-saudaraku, **Arief Hidayat, Nurul Mukhoromah, Ikhsan Nur Karim**. Terima kasih atas segala kasih sayang,kesabaran, doa, dan dukungan yang besar yang diberikan kepada penulis sehingga dapat menyelesaikan skripsi ini dengan baik.
3. Spesial untuk yang tersayang telah menemani beberapa tahun ini **Katherine M. Manuputty, A. Ade Asrindah, S.Si.,Egidia Triayu Tulak, S.Si., dan Eka Fahreza Hatta, S.Si**. sahabat yang selama ini mengisi hari-hari penulis selama beberapa tahun terakhir dalam menjalani rutinitas perkuliahan, membantu, memotivasi dan mendukung dalam menyelesaikan skripsi.
4. Teman-teman seperjuangan di Jurusan Statistika terkhusus **Nurwasari, S.Si.,Reski Wahyunik, S.Si., Putri Indi Rahayu, S.Si.,M.Si., Riska Arruan B.S, dan Seluruh Statistika 2013 yang tidak sempat disebutkan.** Terima kasih atas kebersamaannya selama ini. Semoga Allah membalas kebaikan kalian dengan yang lebih baik.

5. Teman-teman **KKN Gel. 93 Kec. Pajukukang, Bantaeng** terkhusus posko **Desa Biangkeke** yang telah menjadi saudara bagi penulis dan semoga kedepannya silahturahmi kita tetap terjalin.
6. Teman-teman **Husnul Khatimah,S.IP., Kamsinar, S.Si**., **Hena Suri Intan, S.Si.** Terima kasih atas dukungan yang diberikan kepada penulis
7. Keluarga besar **HIMATIKA FMIPA UNHAS**, terkhusus **Binomial 2013**, Penulis merasa bangga menjadi salah satu bagian dari himpunan ini. Persaudaraan kita begitu kuat. **BRAVO HIMATIKA.**
8. Kepada semua pihak yang tidak dapat penulis tuliskan satu persatu. Terima kasih atas dukungan, doa, dan partisipasi yang diberikan dalam penyusunan tugas akhir ini.

Serta kepada seluruh pihak yang tidak dapat penulis sebutkan satu persatu, terima kasih untuk semuanya. Semoga apa yang telah dituliskan oleh penulis pada skripsi ini dapat bermanfaat bagi semuanya. Penulis menyadari bahwa dalam penyusunan skripsi ini masih terdapat kekurangan-kekurangan. Oleh karena itu,penulis mengharapkan saran dan kritik yang bersifat membangun demi skripsi yang lebih baik lagi. Semoga skripsi ini dapat bermanfaat bagi pihak-pihak yang berkepentingan.

Makassar,14 Mei 2020

Penulis

PERNYATAAN PERSETUJUAN PUBLIKASI
TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

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

Nama	:	Fitri Annisa
NIM	:	H 121 13 002
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 Noneksklusif (*Non-exclusive Royalty - Free Right*)** atas karya ilmiah saya yang berjudul:

"PERFORMA MODEL *STATISTICAL DOWNSCALING* DENGAN PEUBAH DUMMY BERDSARKAN K-MEANS DAN AVERAGE LINKAGE"

Beserta perangkat yang ada (jika diperlukan). Terkait dengan hal di atas, maka pihak universitas berhak menyimpan, mengalih-media/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan 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 14 Mei 2020

Yang Menyatakan

(Fitri Annisa)

ABSTRAK

Perubahan iklim merupakan salah satu isu yang cukup ramai dibicarakan belakangan ini. Hal ini disebabkan karena dampak perubahan iklim tersebut sudah sangat dirasakan pada setiap aspek-aspek kehidupan manusia. *Statistical downscaling* (SD) merupakan salah satu alternatif untuk mengatasi masalah tersebut. *Statistical downscaling* adalah model statistika yang dapat menghubungkan perubahan iklim yang berskala global (presipitasi) dengan peubah iklim yang berskala lokal (curah hujan). Data GCM yang berskala besar (global) mengandung multikolinearitas. Metode yang digunakan dalam penilitian ini adalah estimasi parameter model regresi liu dan regresi komponen utama untuk mengatasi masalah multikolinearitas. Terdapat 4 komponen utama (KU) yang optimal untuk digunakan pada model SD dengan regresi komponen utama. Selain itu, peubah *dummy* berdasarkan *k-means* dan *Average Linkage* digunakan dalam model untuk mengatasi ragam sisaan yang heterogen. Hasil penelitian curah hujan lokal yang diperoleh estimasi parameter model regresi liu *dummy* (k-means) merupakan model yang terbaik berdasarkan nilai koefisien determinasi (R^2) 85,03% dan korelasi yang tinggi yaitu 0,955 dengan *root mean square error* (RMSE) 77.83 dan *root mean square error* prediction *(*RMSEP) 122.09 yang lebih rendah. Sedangkan regresi komponen utama dummy (R^2) 95,80% dan nilai RMSE 63.46 dengan nilai korelasi tertinggi yaitu 0,828 dan RMSEP 438.14.

Kata Kunci: *Average Linkage,* Estimasi Parameter Model Regresi Liu, *Global Circulation Model, K-Means,* Multikolinearitas, Peubah *dummy*, Regresi Komponen Utama, *Statistical Downscalling*.

ABSTRACT

Climate change is one of the issues that has been discussed quite recently. This is because the impact of climate change is already being felt in every aspect of human life. Statistical downscaling (SD) is an alternative to solve this problem. Statistical downscaling is statistical model that can connect the global circulation model (GCM) output climate variable (precipitation) with the local scale climate variable (rainfall). GCM data which on a large scale contains multicollinearity. The method used in this research is estimation liu regression and principal analysis regression to overcome multicollinearitas. There are 4 principal components (KU) which is optimal for use of the SD model with principal component regression. In addition, dummy variables based on k-means and Average Linkage are used in the model to overcome heterogeneous variance. The results of local rainfall research obtained by estimating the parameters of the liu dummy regression model (k-means) is the best model based of the coefficient of determination (R^2) 85.03% and a high correlation of 0.955 with root mean square error (RMSE) 77.83 and 122.09 lower root mean square error prediction (RMSEP). While the principal component regression dummy (R^2) is 95.80% and the RMSE value is 63.46 with the highest correlation value which is 0.828 and RMSEP 438.14.

Keywords: Average Linkage, Estimastion Liu Regression, Global Circulation Model, K-Means, Dummy Variabel, Principal Component Regression, Statistical Downscaling.

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR LAMPIRAN

BAB I
PENDAHULUAN

1.1. Latar Belakang

Indonesia mempunyai garis pantai 95.181 km dan luas lautnya sekitar 5,8 juta km^2 atau 70% dari luas seluruh Indonesia. Berbagai sektor tercakup di dalamnya, mulai dari masyarakat pesisirnya, nelayan, pulau-pulau kecil, perikanan, sampai sumber daya kelautan lainnya termasuk salah satunya garam. Produksi garam merupakan salah satu isu nasional yang menjadi perhatian pemerintah saat ini (Haidawati, 2014).

Kabupaten Pangkep, Provinsi Sulawesi Selatan merupakan salah satu daerah penghasil garam terbaik di Indonesia dan terpusat di beberapa tempat seperti di daerah Bonto mania dengan hasil 151,1 ton garam dan Maccini baji 21,7 ton garam. Ini merupakan suatu kekayaan tersendiri bagi daerah Kabupaten Pangkep (Alfiandhi, 2012). Pola pertanian garam yang digunakan dalam masyarakat di Kabupaten Pangkep masih menggunakan cara tradisonal. Kendala-kendala yang dihadapi oleh petambak garam yaitu penurunan produksi yang diakibatkan oleh faktor cuaca yang tidak menentu (Haidawati, 2014).

Perubahan iklim yang terjadi sering digunakan untuk meramalkan keadaan iklim yang akan datang. Untuk pendugaan iklim yang akan datang hanya mungkin dilakukan dengan menggunakan model iklim. Salah satu model iklim yang digunakan untuk meramalkan keadaan iklim adalah *Global Circulation Models* (GCM). GCM adalah alat terpenting dalam upaya memahami sistem iklim. GCM mengahasilkan data dalam bentuk *grid* atau petak wilayah dengan resolusi rendah ($2{,}5^{\circ} atau\ \pm 300\ km^2$) yang merepresentasikan keadaan iklim global tapi tidak dalam skla regional atau skala lokal. Untuk menggunakan data GCM sebagai penduga bagi penelitian perubahan iklim local, diperlukan pendekatan yang mampu mengatasi perubahan iklim tersebut. Salah satu pendekatan untuk mengatasi masalah tersebut digunakan Teknik *Downscaling* (Effendi, 2009).

Teknik *downscaling* adalah suatu proses transformasi data dari suatu grid dengan unit skala besar menjadi data pada grid-grid dengan unit skala yang lebih kecil. Wilby & Wigley (1997) menyatakan bahwa *downscaling* adalah suatu cara menginterpolasi variable-variabel predictor atmosfir berskala regional terhadap variable-variabel berskala lebih kecil. Salah satu jenis teknik *downscaling* adalah *statistical downscaling* (SD), di mana merupakan model statistik yang menggambarkan hubungan antara data pada grid-grid berskala besar (GCM) dengan data pada grid berskala lebih kecil. Data GCM yang berskala besar memungkinkan adanya multikolinearitas (Purnomoadi,2009).

Multikolinearitas menyebabkan standar eror penduga parameter regresi menjadi besar dan selang kepercayaan menjadi lebar sehingga penggunaan metode kuadrat terkecil (MKT) menjadi tidak valid (Montgomery & Peck, 1992). Regresi komponen utama (RKU) merupakan salah satu metode yang paling sering digunakan dalam pemodelan SD. Metode lain yang serupa dengan metode RKU dalam mengatasi multikolinearitas yaitu metode estimasi liu. Pada tahun 1993 Liu memperkenalkan suatu metode yang dapat mengatasi multikolinearitas, metode tersebut disebut metode estimasi liu. Seperti halnya regresi ridge, metode estimasi liu juga memiliki nilai konstantataksiran yang digunakan untuk mengatasi multikolinearitas.

Pemodelan SD yang telah dilakukan sebelumnya diantaranya Sahriman (2014) menggunakan regresi kuadrat terkecil parsial (RKTP) dan regresi komponen utama (RKU). Dalam penelitiannya, dilakukan penambahan variabel *dummy* untuk mengatasi kondisi sisaan ragam yang tidak homogen pada pemodelan SD.

Analisis cluster termasuk teknik multivariat yang bertujuan untuk mengelompokkan objek-objek berdasarkan kesamaan karakteristik di antara objek-objek tersebut. Prinsip dari teknik Kluster adalah homogenitas (kesamaan) yang tinggi antar kelompok dalam satu Kluster dan heterogenitas (perbedaan) yang tinggi antar Kluster yang satu dengan *cluster* yang lainnya. Secara umum terdapat dua teknik Kluster, yakni analisis Kluster hierarki dan non-hierarki (Mattjik & Sumertajaya, 2011).

Teknik kluster hirarki adalah K-*means*. *K-Means* merupakan suatu algoritma pengklasteran yang cukup sederhana yang mempartisi databest kedalam beberapa clasteran *k.Algoritma* cukup mudah untuk diimplementasikan dan dijalakan, relatif cepat, mudah disesuaikan dan banyak digunakan (Wu & Kumar, 2009). Sedangkan salah satu teknik kluster non-hierarki adalah *Average Linkage*. *Average Linkage* adalah metode pengklasteran yang didasarkan pada jarak rata-rata antar objeknya. Objek yang mempunyai jarak terdekat bergabung dan membentuk klaster baru(Rachmatin,2014).

Berdasarkan uraian latar belakang tersebut, penulis mengambil judul " **Performa Model *Statistical Downscaling* dengan Peubah *Dummy* Berdasarkan Metode *K-Means* dan *Average Linkage***".

1.2. Rumusan Masalah

Berdasarkan latar belakang masalah, disusun rumusan masalah sebagai berikut

a. Bagaimana cara penentuan peubah *dummy k-means* dan *average linkage* dalam model *statistical downscaling* ?
b. Bagaiamana cara memperoleh model estimasi parameter model regresi liu dan model RKU dalam penentuan peubah *dummy* k-*means* dan *average linkage* dengan model *statistical downscalling* pada data curah hujan?

1.3. Batasan Masalah

Masalah dalam penulisan ini dibatasi hanya difokuskan pada :

1. Penentuan peubah *dummy* k-*means* dan *average linkage* pada model *statistical dwonscalling.*
2. Model *statistical downscaling* dalam penentuan peubah *dummy k-means* dan *average linkage.* Penanganan Multikolinearitas dengan menggunakan metode estimasi parameter model regresi liu dan regresi komponen utama untuk data curah hujan tahun 1981-2016 dengan kriteria pemilihan komponen utama dengan proporsi kumulatif keragaman minimal 80%.

1.4. Tujuan Penelitian

Tujuan penelitian berdasarkan rumusan masalah adalah

1. Memperoleh peubah *dummy* k-*means* dan *average linkage* dengan model *statistical downscaling* pada data curah hujan.

2. Memperoleh model estimasi parameter model regresi liu dan model regresi komponen utama dengan peubah *dummy* berdasarkan k-*means* dan *average linkage* dalam model *statistical downscaling*.

1.5. Manfaat Penelitian

Manfaat dari penelitian ini adalah sebagai referensi kepada pembaca tentang cara peramalan curah hujan menggunakan estimasi parameter model regresi liu dan regresi komponen utama berdasarkan *k-means* dan *average linkage* dalam model *statistical downscaling* dan juga sebagai sumber informasi tentang curah hujan di daerah Kabupaten Pangkep

BAB II
TINJUAN PUSTAKA

2.1.Global Circulation Model

Global circulation model (GCM) merupakan model matematis yang menggambarkan sirkulasi umum atmosfer di bumi. Model ini menduga perubahan unsur-unsur cuaca dalam bentuk luaran grid-grid yang berukuran 100-500 km menurut lintang dan bujur (Sutikno, 2008). GCM merupakan suatu alat penting dalam studi keragaman iklim dan perubahan iklim (Zorita dan Storch, 1999). Namun informasi GCM masih berskala global, sehingga sulit untuk memperoleh langsung informasi berskala lokal dari GCM. Tetapi GCM masih mungkin digunakan untuk memperoleh informasi skala local atau regional bila teknik *downscaling* digunakan (Fernandez 2005).

Downscaling adalah suatu metode untuk mendapatkan informasi tentang perubahan iklim atau iklim beresolusi tinggi dari model iklim global yang relatif kasar, dimana model iklim global memiliki resolusi 150-300 km. Akan tetapi, banyak dampak model memerlukan informasi pada skala 50 km atau kurang, sehingga beberapa metode dibutuhkan untuk memperkirakan informasi dengan skala kecil. Untuk menjembatani skala GCM yang besar dengan skala yang lebih kecil (kawasan yang menjadi studi) digunakan teknik *Statistical Downscaling* (Lembang, 2013)

2.2.Statistical Downscalling

Statistical downscaling (SD) adalah pendekatan empiris mengenai hubungan secara statistika antara atmosfer global (GCM) dengan curah hujan. Ide dasar dari SD adalah menentukan parameter hubungan antara iklim skala global dengan iklim skala lokal dan menggunakan hubungan ini untuk proyeksi hasil simulasi GCM pada iklim masa lalu, sekarang, atau masa depan yang berskala lokal. SD menggunakan model statistik dalam menggambarkan hubungan antara data pada grid berskala global (prediktor) dengan data pada grid yang berskala lokal (respon) untuk menterjemahkan anomali-anomali skala global menjadi anomali dari beberapa peubah iklim lokal (Zorita & von Storch, 1999)

Bentuk umum model SD dapat disajikan pada persamaan :

$$y_{t\times 1} = f(X_{t\times p}) \tag{2.1}$$

Dengan :

$y_{t\times 1}$: peubah tak bebas (curah hujan lokal (mm/bulan))

$X_{t\times p}$: peubah bebas (curah hujan permukaan GCM (mm/bulan))

t : banyaknya waktu (misal bulanan)

p : banyaknya grid domain GCM

Pada umumnya model *SD* melibatkan data deret waktu (*t*) dan data spasial GCM (*p*). Kompleksitas model ini terjadi karena **X** berkorelasi dan pengamatan peubah **y** berotokorelasi (Wigena, 2006).

2.3.Multikolinearitas

Istilah multikolinieritas mula-mula ditemukan oleh Ragnar Frisch pada tahun 1934 yang berarti adanya hubungan linier antara variabel (x_j). Maksud dari adanya hubungan linier antara variabel adalah sebagai berikut: misalkan hubungan linier antara x_1 dan x_2. Jika model ini diestimasi dengan metode kuadrat terkecil, β dapat ditentukan, tetapi variansi yang dihasilkan besar yang mengakibatkan erornya besar dan interval kepercayaannya semakin besar, sehinggga kurang tepat. Permasalahan ini mengakibatkan dampak yang tidak baik bagi model. Pada analisis regresi, multikolinieritas dikatakan ada apabila beberapa kondisi berikut dipenuhi:

a. Dua variabel berkorelasi sempurna (oleh karena itu vektor-vektor yang menggambarkan variabel tersebut adalah kolinier).

b. Dua variabel prediktor hampir berkorelasi sempurna yaitu koefisien korelasinya mendekati 1.

c. Kombinasi linier dari beberapa variabel prediktor berkorelasi sempurna atau mendekati sempurna dengan variabel prediktor yang lain.

d. Kombinasi linier dari satu sub-himpunan variabel prediktor berkorelasi sempurna dengan satu kombinasi linier dari sub-himpunan variabel prediktor yang lain.

Cara yang digunakan untuk menguji suatu data yang mengalami masalah multikolinearitas yaitu (Myers, 1990) :

1. Menentukan matriks korelasi dari semua variabel independen

   Prosedur ini merupakan pemeriksaan yang paling sederhana dan paling mudah. Nilai korelasi yang tinggi antara variabel satu dengan yang lainnya memperlihatkan adanya multikolinearitas. Jika didapati elemen-elemen matriks korelasi yang besar yaitu mendekati 1 atau -1, maka dicurigai terdapat masalah multikolinearitas pada data.

2. *Variance Inflation Factor*

   Nilai *Variance Inflation Factor* (*VIF*) juga dapat digunakan untuk memeriksa adanya multikolinearitas. Apabila salah satu dari nilai *VIF* lebih dari 10 maka dapat diidentifikasikan bahwa dalam variabel independen terdapat masalah multikolinearitas. Nilai *VIF* dapat diperoleh dengan persamaan:

$$VIF = \frac{1}{1-R_j{}^2} \tag{2.2}$$

dimana R_j^2 adalah koefisien determinasi dari x_j dan x_j^* variabel bebas lainnya. Nilai R_j^2 memiliki pengaruh terhadap nilai *VIF* yang dihasilkan, dimana semakin besar nilai R_j^2 maka semakin besar pula nilai *VIF* yang dihasilkan. Jika nilai $VIF > 10$ maka dapat disimpulkan bahwa terdapat masalah multikolinieritas pada variabel tersebut. Dengan melihat nilai VIF itu juga dapat membantu kita dalam mengidentifikasi variabel-variabel mana yang mengalami masalah multikolinieritas (Gujarati, 2004).

2.4.Analisis Regresi Berganda

Analisis regresi merupakan analisis statistika yang bermanfaat untuk mengetahui hubungan antara dua variabel atau lebih sehingga salah satu variabel dapat diduga dari variabel lainnya. Dalam analisis regresi ini dapat diketahui bentuk atau pola hubungan yang terjadi serta dapat dilakukan prediksi berdasarkan nilai variabel yang sudah diketahui. Jika analisis regresi dilakukan untuk satu varibel tak bebas (y) dengan lebih dari satu variabel bebas (X) maka regresi ini dinamakan regresi linier berganda(Montgomery & Peek, 1992).

Secara umum persamaan regresi linier berganda dengan p variabel bebas dinyatakan dengan :

$$y_i = \beta_0 + \sum_{j=1}^{p} \beta_j X_{ij} + \varepsilon_i \qquad j = 1,2,3\ldots,p$$

$$i = 1,2,3 \ldots, t \qquad (2.3)$$

dalam bentuk matriks dinyatakan sebagai berikut:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \qquad (2.4)$$

dengan :

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_t \end{bmatrix}, \boldsymbol{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{t1} & x_{t2} & \cdots & x_{tp} \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}, \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_t \end{bmatrix}$$

dimana:

$\boldsymbol{y}$ adalah vektor pada variabel tak bebas yang berukuran $(t \times 1)$

$\boldsymbol{X}$ adalah matriks dari p variabel bebas yang berukuran $(t \times (p+1))$

$\boldsymbol{\beta}$ adalah vektor parameter regresi yang berukuran $((p+1) \times 1)$

$\boldsymbol{\varepsilon}$ adalah vektor sisaan yang berukuran $(t \times 1)$

2.5.Nilai Eigen dan Vektor Eigen

Kata "eigen" dalam bahasa Jerman berarti "tepat". Secara historis nilai eigen juga disebut nilai yang tepat, nilai karakteristik, atau akar laten, dan terminologi yang sama berlaku untuk vektor eigen (Richard,1986).

Definisi:

Misalkan A matrik berordo $n \times n$, vektor $x \in R^n$ dan $x \neq 0$ disebut vektor eigen, jika terdapat bilangan riil λ, yang disebut nilai eigen, sehingga memenuhi persamaan:

$$\boldsymbol{Ax} = \lambda \boldsymbol{x}$$

Dari definisi di atas dapat diketahui persyaratan-persyaratan untuk nilai eigen maupun vektor eigen. Nilai eigen λ merupakan bilangan riil, yang berarti dapat bernilai nol, negatif dan juga positif, sedangkan vektor eigen **x** merupakan anggota dari R^n untuk $\boldsymbol{A}_{nxn}$ dan **x** *bukan* vektor nol (Anton,1997).

Mencari nilai eigen dari matriks A yang berordo $n \times n$, dapat kita mulai dengan menuliskan persamaan (Richard,1986):

$$\boldsymbol{Ax} = \lambda \boldsymbol{x}$$

$$\boldsymbol{Ax} = \lambda \boldsymbol{Ix}$$

$$\lambda \boldsymbol{Ix} - \boldsymbol{Ax} = 0$$

$$(\lambda \boldsymbol{I} - \boldsymbol{A})\boldsymbol{x} = 0 \qquad (2.5)$$

Dapat kita lihat λ adalah nilai eigen dari $\boldsymbol{A}$ sehingga $(\lambda \boldsymbol{I} - \boldsymbol{A})\boldsymbol{x} = 0$ adalah SPL dengan solusi sejati (*nontrivial solution*) dan $\lambda I - A$ adalah singular maka:

$$det(\lambda \boldsymbol{I} - \boldsymbol{A}) = 0$$

Polinom λ berderajat n dari persamaan diatas disebut karakteristik polinom dan persamaan diatas merupakan persamaan karakteristik dari matriks **A**.

2.6.Matriks Data Multivariat

Matriks data multivariat disusun dari sampel data multivariate dan dapat ditulis dalam bentuk (Johnson dan Wichern, 2007) berikut :

$$\boldsymbol{X}_{t\times p} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1j} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2j} & \dots & x_{2p} \\ \vdots & \vdots & \dots & \vdots & \vdots & \vdots \\ x_{i1} & x_{i2} & \dots & x_{ij} & \dots & x_{ip} \\ \vdots & \vdots & \dots & \vdots & \vdots & \vdots \\ x_{t1} & x_{t2} & \dots & x_{tj} & \dots & x_{tp} \end{bmatrix} \quad (2.6)$$

dimana:

x_{ij} : pengamatan antara variabel ke-i dengan variabel ke-j

t : jumlah observasi

p : jumlah variabel

2.7.Matriks Kovariansi dan Korelasi

Matriks kovariansi(matriks varian-kovarian) disimbolkan dengan Σ. Matriks kovariansi dapat diperoleh dengan cara (Johnson dan Wichern, 2007) sebagai berikut:

$$E(\boldsymbol{X}) = \begin{bmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_P) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_P \end{bmatrix} = \boldsymbol{\mu} \quad (2.7)$$

Cov($\boldsymbol{X}$) = $\Sigma = E(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})'$

$$= E\left(\begin{bmatrix} X_1-\mu_1 \\ X_2-\mu_2 \\ \vdots \\ X_P-\mu_P \end{bmatrix} [X_1-\mu_1, X_2-\mu_2, \dots, X_P-\mu_P]\right)$$

$$\text{Cov}(\boldsymbol{X}) = E\begin{bmatrix} (X_1-\mu_1)^2 & (X_1-\mu_1)(X_2-\mu_2) & \dots & (X_1-\mu_1)(X_P-\mu_P) \\ (X_2-\mu_2)(X_1-\mu_1) & (X_2-\mu_2)^2 & \dots & (X_2-\mu_2)(X_P-\mu_P) \\ \vdots & \vdots & \ddots & \vdots \\ (X_p-\mu_p)(X_1-\mu_1) & (X_p-\mu_p)(X_2-\mu_2) & \dots & (X_P-\mu_P)^2 \end{bmatrix}$$

$$= \begin{bmatrix} E(X_1-\mu_1)^2 & E(X_1-\mu_1)(X_2-\mu_2) & \dots & E(X_1-\mu_1)(X_P-\mu_P) \\ E(X_2-\mu_2)(X_1-\mu_1) & E(X_2-\mu_2)^2 & \dots & E(X_2-\mu_2)(X_P-\mu_P) \\ \vdots & \vdots & \ddots & \vdots \\ E(X_p-\mu_p)(X_1-\mu_1) & E(X_p-\mu_p)(X_2-\mu_2) & \dots & E(X_P-\mu_P)^2 \end{bmatrix}$$

$$\mathbf{\Sigma}_{\mathrm{p\times p}} = \mathrm{cov}(\mathbf{X}) = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix} \tag{2.8}$$

Ukuran keeratan linear antara variabel random X_i dan X_j adalah koefisien korelasi populasi $\rho_{i,j}$ yang dinyatakan dalam varians σ_{ii} dan kovarians σ_{ij}, yaitu :

$$\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}}\sqrt{\sigma_{jj}}} \tag{2.9}$$

Matriks koefisien korelasi adalah matriks simetri berukuran pxp, yaitu :

$$\boldsymbol{\rho}_{\boldsymbol{p\times p}} = \begin{bmatrix} \frac{\sigma_{11}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{11}}} & \frac{\sigma_{12}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}} & \cdots & \frac{\sigma_{1p}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{pp}}} \\ \frac{\sigma_{12}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{11}}} & \frac{\sigma_{22}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{22}}} & \cdots & \frac{\sigma_{2p}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{pp}}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\sigma_{1p}}{\sqrt{\sigma_{pp}}\sqrt{\sigma_{11}}} & \frac{\sigma_{2p}}{\sqrt{\sigma_{pp}}\sqrt{\sigma_{22}}} & \cdots & \frac{\sigma_{pp}}{\sqrt{\sigma_{pp}}\sqrt{\sigma_{pp}}} \end{bmatrix}$$

$$\boldsymbol{\rho}_{\boldsymbol{p\times p}} = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{12} & 1 & \cdots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1p} & \rho_{2p} & \cdots & 1 \end{bmatrix} \tag{2.10}$$

2.8.Matriks Orthogonal

Sebuah matriks nonsingular **Q** disebut orthogonal jika dan hanya jika $\mathbf{Q}^{t}=\mathbf{Q}^{-1}$. Jika **A** dan **B** adalah sebuah matriks orthogonal dengan ukuran yang sama maka dituliskan $|\boldsymbol{A}| = \pm\mathbf{1}$, $\boldsymbol{A}^{\boldsymbol{t}} = \boldsymbol{A}^{-\mathbf{1}}$ adalah orthogonal dan $\boldsymbol{AB}$ adalah orthogonal (Andrilli, 2016).

Sebuah Matriks bujursangkar **A** dapat dikatakan terdiagonalisasi secara ortogonal bila terdapat matriks **Q** ortogonal sehingga $\mathbf{Q}^{-1}\mathbf{AQ}$ merupakan matriks diagonal. Matriks **Q** dikatakan mendiagonalisasikan **A** secara ortogonal.

Dimisalkan $\boldsymbol{B} = \mathbf{Q}^{-1}\mathbf{AQ}$

maka $\quad \mathbf{Q}\boldsymbol{B}\mathbf{Q}^{-1} = \mathbf{A}$

$\mathbf{Q}\boldsymbol{B}\mathbf{Q}^{\mathrm{t}} = \mathbf{A}$ (sifat $\mathbf{Q}^{-1} = \mathbf{Q}^{\mathrm{t}}$ matriks ortogonal)

$[\mathbf{Q}\boldsymbol{B}\mathbf{Q}^{\mathrm{t}}]^{\mathrm{t}} = \mathbf{A}^{\mathrm{t}}$ (ruas kiri dan ruas kanan ditrasposekan)

$\mathbf{Q}\boldsymbol{B}^{\mathrm{t}}\mathbf{Q}^{\mathrm{t}} = \mathbf{A}^{\mathrm{t}}$ (sifat transpose dari perkalian dua matriks)

$\mathbf{Q}\boldsymbol{B}\mathbf{Q}^{\mathrm{t}} = \mathbf{A}^{\mathrm{t}}$ (matriks diagonal, $\boldsymbol{B} = \boldsymbol{B}^{t}$)

$\mathbf{A} = \mathbf{A}^{\mathrm{t}}$.

Dapat kita simpulkan bahwa matriks bujur sangkar dapat didiagonalisasi apabila matriks tersebut merupakan matriks simetris.

2.9. Metode Kuadrat Terkecil

Metode kuadrat terkecil (MKT) merupakan suatu metode yang digunakan untuk mengestimasi parameter regresi dalam analisis regresi berganda. Metode ini bertujuan untuk meminimunkan jumlah kuadarat galat (Montomery dan Peck, 1992). Misalkan model yang akan diestimasi adalah parameter dari Persamaan (2.5) yaitu:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

Untuk mendapatkan nilai penduga atau estimasi dari $\boldsymbol{y}$ yaitu $\widehat{\boldsymbol{y}} = \boldsymbol{X}\widehat{\boldsymbol{\beta}}$ diperlukan nilai estimator untuk parameter $\boldsymbol{\beta}$ yaitu $\widehat{\boldsymbol{\beta}}$. Estimasi regersinya adalah $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_n$ dan dalam bentuk vektor yang dilambangkan sebagai berikut:

$$\widehat{\boldsymbol{\beta}}_{(p+1)\times 1} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_p \end{bmatrix}$$

untuk mengestimasi parameter regresi menggunakan MKT, yaitu dengan cara meminumkan jumlah kuadrat galat $\sum_{i=1}^{n} \varepsilon_i{}^2$. Misalakan fungsi jumlah kuadrat galat yaitu:

$$S(\boldsymbol{\beta_0}, \boldsymbol{\beta_1}, \dots, \boldsymbol{\beta_p}) = \sum_{\boldsymbol{i=1}}^{n} \boldsymbol{\varepsilon_i}^2 = \sum_{i=1}^{n} \left(\boldsymbol{y_i} - \boldsymbol{\beta_0} - \sum_{\boldsymbol{j=1}}^{p} \boldsymbol{\beta_j X_{ij}}\right)^2 \qquad (2.11)$$

Bentuk matriks dari Persamaan (2.6) dapat ditulis

$$S(\boldsymbol{\beta}) = \sum_{i=1}^{n} \boldsymbol{\varepsilon}_i{}^2 = \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} = (\boldsymbol{y} - \boldsymbol{X\beta})'(\boldsymbol{y} - \boldsymbol{X\beta})$$

$$= \boldsymbol{y}'\boldsymbol{y} - \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{y} - \boldsymbol{y}'\boldsymbol{X\beta} + \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X\beta} \qquad (2.12)$$

Persamaan (2.7) diturunkan terhadap $\boldsymbol{\beta}$, sehingga didapat nilai estimasi parameter $\widehat{\boldsymbol{\beta}}$ yaitu:

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-\mathbf{1}}\boldsymbol{X}'\boldsymbol{y} \qquad (2.13)$$

Nilai *Mean Square Error* (MSE) metode kuadrat terkecil dapat di cari dengan rumus berikut (Adegoke *et al.* 2016):

$$\mathrm{MSE}_{\mathrm{MKT}} = \hat{\sigma}^2 \sum_{i=1}^{p} \frac{1}{\lambda_i} \qquad (2.14)$$

dengan;

$\hat{\sigma}^2 =$ variansi MKT

$\lambda_i =$ nilai eigen dari matriks $(\mathbf{X}'\mathbf{X})$

2.10. Estimasi Parameter Model Regresi LIU

Mengingat $\boldsymbol{X}'\boldsymbol{X}$ merupakan matriks simetris, sehingga terdapat matriks orthogonal , sedemikian hingga:

$$\mathbf{Q}'(\mathbf{X}'\mathbf{X})\mathbf{Q} = \boldsymbol{\Lambda}$$

$$\mathbf{Q}'\mathbf{X}'\mathbf{X}\mathbf{Q} = \boldsymbol{\Lambda}$$

$$(\mathbf{XQ})'(\mathbf{XQ}) = \boldsymbol{\Lambda}$$

$$\mathbf{Z}'\mathbf{Z} = \boldsymbol{\Lambda}$$

Dengan $\boldsymbol{\Lambda}$ merupakan matriks yang berukuran $p \times p$ dengan anggota dari diagonal utamanya merupakan nilai eigen $(\lambda_1, \lambda_2, \cdots, \lambda_{\mathrm{p}})$ dari matriks $\boldsymbol{X}'\boldsymbol{X}$ atau dapat ditulis $\boldsymbol{\Lambda} = diag(\lambda_1, \lambda_2, \cdots, \lambda_{\mathrm{p}})$ dan matriks $\mathbf{Q}$ adalah matriks orthogonal yang berukuran $p \times p$ yang elemen-elemennya adalah nilai eigen vektor dari $\boldsymbol{X}'\boldsymbol{X}$.

$$\mathbf{X}'\mathbf{X} = \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}' \quad \text{dan} \quad \mathbf{Q}'\mathbf{Q} = \mathbf{Q}\mathbf{Q}' = \mathbf{I}$$

Sehingga model regresi pada Persamaan (2.5) dapat ditulis sebagai berikut:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{Q}\boldsymbol{Q}'\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$$\boldsymbol{y} = (\boldsymbol{X}\boldsymbol{Q})(\boldsymbol{Q}'\boldsymbol{\beta}) + \boldsymbol{\varepsilon}$$

$$\boldsymbol{y} = \boldsymbol{Z}\boldsymbol{\alpha} + \boldsymbol{\varepsilon} \tag{2.15}$$

dengan

$\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{Q}$ dan $\boldsymbol{\alpha} = \mathbf{Q}'\boldsymbol{\beta}$

dengan menggunakan MKT dan berdasarkan Persamaan (2.15) dengan estimasi $\hat{\boldsymbol{\alpha}}_{\mathrm{MKT}}$ adalah

$$\hat{\boldsymbol{\alpha}}_{\mathrm{MKT}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}$$

$$\hat{\boldsymbol{\alpha}}_{\mathrm{MKT}} = \boldsymbol{\Lambda}^{-1}\mathbf{Z}'\mathbf{y}$$

dari Persamaan diatas dapat dibentuk menjadi:

$$\hat{\boldsymbol{\alpha}} = ((\boldsymbol{X}\boldsymbol{Q})'(\boldsymbol{X}\boldsymbol{Q}))^{-1}(\boldsymbol{X}\boldsymbol{Q})'\mathbf{y}$$

$$\hat{\boldsymbol{\alpha}} = (\boldsymbol{Q}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{Q})^{-1}\mathbf{Q}'\mathbf{X}'\,\mathbf{y}$$

$$\hat{\boldsymbol{\alpha}} = (\boldsymbol{Q}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{Q})^{-1}\mathbf{Q}'\mathbf{X}'\,\mathbf{X}\hat{\boldsymbol{\beta}}$$

$$\hat{\boldsymbol{\alpha}} = (\boldsymbol{Q}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{Q})^{-1}\mathbf{Q}'\mathbf{X}'\,\boldsymbol{X}\boldsymbol{Q}\boldsymbol{Q}'\,\hat{\boldsymbol{\beta}}$$

$$\hat{\boldsymbol{\alpha}} = (\boldsymbol{Q}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{Q})^{-1}(\mathbf{Q}'\mathbf{X}'\boldsymbol{X}\boldsymbol{Q})\,\boldsymbol{Q}'\,\hat{\boldsymbol{\beta}}$$

$$\hat{\boldsymbol{\alpha}} = \boldsymbol{Q}'\,\hat{\boldsymbol{\beta}}$$

atau

$$\hat{\boldsymbol{\beta}} = \mathbf{Q}\,\hat{\boldsymbol{\alpha}}$$

Dalam mengatasi masalah multikolinearitas, Kejian Liu (1993) mengusulkan metode *Liu Estimator* yaitu suatu metode yang dapat mengatasi masalah multikolinearitas, dengan menambahkan augmentasi $d\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + \tilde{\boldsymbol{\varepsilon}}$ ke

Persamaan (2.4) dan dengan menggunakan metode MKT maka akan diperoleh estimasi LIU yang dirumuskan sebagai berikut (Liu, 1993):

$$\widehat{\boldsymbol{\beta}}_{LE} = (\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{I})^{-1}(\boldsymbol{X}'\boldsymbol{X} + d_L\boldsymbol{I})\widehat{\boldsymbol{\beta}}_{\text{MKT}} \tag{2.16}$$

atau

$$\widehat{\boldsymbol{\alpha}}_{LE} = (\boldsymbol{\Lambda} + \boldsymbol{I})^{-1}(\boldsymbol{\Lambda} + d_L\boldsymbol{I_P})\widehat{\boldsymbol{\alpha}}_{\text{MKT}}$$

Dimana d adalah konstanta *LIU* dan dirumuskan sebagai berikut (Liu, 1993):

$$d_L = \frac{\sum_{j=1}^{p}\left[\dfrac{\hat{\alpha}_{MKT_j}^2 - \hat{\sigma}_{MKT}^2}{\left(\lambda_j + 1\right)^2}\right]}{\sum_{j=1}^{p}\left[\dfrac{\lambda_j\hat{\alpha}_{MKT_j}^2 + \hat{\sigma}_{MKT}^2}{\lambda_j\left(\lambda_j + 1\right)}\right]} \tag{2.17}$$

dengan $\widehat{\boldsymbol{\alpha}} = \boldsymbol{Q}'\widehat{\boldsymbol{\beta}}_{\text{MKT}}$ dan $\hat{\sigma}_{\text{MKT}}^2 = \frac{(y-\hat{y})\prime(y-\hat{y})}{n-p}$ adalah estimasi MKT dari α dan variansi eror dari MKT, dan $\mathbf{Q}$ adalah matriks orthogonal yang berukuran $p \times p$ yang elemen-elemennya adalah nilai eigen vektor dari $\boldsymbol{X}'\boldsymbol{X}$, dengan nilai d $0 < d < 1$.

Nilai *Mean Square Error* (MSE) estimasi *LIU* dapat dicari dengan menggunakan rumus berikut (imdadullah *et al.* 2017):

$$MSE_{LIU} = \sigma^2 \sum_{i=1}^{p} \frac{(\lambda_i + d)^2}{\lambda_i(\lambda_i+1)^2} + (d-1)^2 \sum_{i=1}^{p} \frac{\hat{\alpha}^2}{(\lambda_i+1)^2} \tag{2.18}$$

2.11. Regresi Komponen Utama

Regresi Komponen Utama (RKU) adalah salah satu teknik analisis multivariat yang bertujuan mereduksi dimensi data di mana sejumlah besar peubah saling terkait satu dengan yang lain, dengan mempertahankan sebesar mungkin keragaman data menjadi kumpulan data baru yang tidak berkorelasi lagi (Joliffe, 2002). RKU pertama kali diperkenalkan oleh Karl Pearson pada tahun 1901. Selanjutnya dikembangkan oleh Hotelling pada tahun 1993 dan Rao pada tahun 1964 hingga saat ini pengembangan RKU masih dilakukan.

Regresi Komponen Utama (RKU) adalah suatu teknik analisis statistik untuk mentransformasi peubah-peubah asli yang masih saling berkorelasi satu dengan yang lain menjadi satu kumpulan peubah baru yang tidak berkorelasi lagi.

Secara umum tujuan dari RKU adalah mereduksi dimensi data. RKU muncul sebagai solusi bagi proses pengumpulan data dimana data tersebut terdiri dari variabel-variabel yang jumlahnya sangat banyak sehingga diperoleh variabel-variabel baru yang jumlahnya lebih sedikit tetapi tetap mampu menjelaskan keragaman data (Johnson dan Wichern, 2007).

Misalkan suatu data terdiri dari n observasi dan p variabel dapat dinyatakan dengan matriks **X** sebagai berikut:

$$\boldsymbol{X}_{t\times p} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{t1} & x_{t2} & \cdots & x_{tp} \end{bmatrix} \qquad (2.19)$$

Persamaan (2.11) dapat dinyatakan dalam bentuk vektor yaitu $\boldsymbol{X} = [\boldsymbol{X_1}, \boldsymbol{X_2}, \dots, \boldsymbol{X_p}]$ dengan $\boldsymbol{X}'_{\boldsymbol{t}} = [\boldsymbol{X}_{t1}, \boldsymbol{X}_{t2}, \dots, \boldsymbol{X}_{tp}]$ dengan $i = 1,2,..,t$, matriks kovarian $\sum$ dan nilai eigen $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$ dengan vektor eigen yang bersesuaian $\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_p$.

Komponen utama yang terbentuk dapat dituliskan sebagai kombinasi linear dari variabel-variabel asalnya yaitu sebagai berikut:

$$\begin{aligned} K_1 &= \boldsymbol{v'_1 X} = v_{11}X_1 + v_{12}X_2 + \cdots + v_{1p}X_p \\ K_2 &= \boldsymbol{v'_2 X} = v_{21}X_1 + v_{22}X_2 + \cdots + v_{2p}X_p \\ &\vdots \qquad\qquad (2.20) \\ K_p &= \boldsymbol{v'_p X} = v_{p1}X_1 + v_{p2}X_2 + \cdots + v_{pp}X_p \end{aligned}$$

atau

$$\boldsymbol{K} = \begin{bmatrix} K_1 \\ K_2 \\ \vdots \\ K_p \end{bmatrix} = \begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1p} \\ v_{21} & v_{22} & \cdots & v_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ v_{p1} & v_{p2} & \cdots & v_{pp} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}$$

dengan

- $var(K_i) = var(\boldsymbol{v_i}'X) = \boldsymbol{v_i}'var(X)\boldsymbol{v_i}$

  sehingga

$var(K_i) = \boldsymbol{v_i}'\sum\boldsymbol{v_i}$; $\quad i = 1,2, \dots , n$ (2.21)

dimana $\Sigma = var(X) = \Sigma_X$

- $cov(K_i, K_k) = cov(\boldsymbol{v_i}'X, \boldsymbol{v_k}'X) = \boldsymbol{v_i}'cov(X, X)\boldsymbol{v_k} = \boldsymbol{v_i}'\sum\boldsymbol{v_k}$

  sehingga

  $cov(K_i, K_k) = \boldsymbol{v_i}'\sum\boldsymbol{v_k}$; $\quad i, k = 1,2, \dots , n$ (2.22)

  dimana $cov(X, X) = var(X) = \Sigma_X$

Langkah-langkah metode RKU adalah sebagai berikut:

- Menghitung matriks varian kovarians Σ dengan menggunakan rumus pada persamaan (2.8)
- Menghitung nilai eigen dari matriks varian kovarians Σ dengan menggunakan rumus sebagai berikut:

$$|\boldsymbol{\Sigma} - \lambda\mathbf{I}| = \mathbf{0} \qquad (2.23)$$

- Menghitung vektor eigen dari matriks varian kovarians Σ dengan menggunakan rumus sebagai berikut:

$$(\boldsymbol{\Sigma} - \lambda\mathbf{I})\boldsymbol{x} = \mathbf{0} \qquad (2.24)$$

- Vektor eigen dinormalisasi dengan menggunakan rumus sebagai berikut:

$$\boldsymbol{e} = \frac{\boldsymbol{x}}{\sqrt{\boldsymbol{x}'\boldsymbol{x}}} \qquad (2.25)$$

- Membentuk komponen utama ke-1,2, … , k dengan menggunakan persamaan (2.20)
- Menghitung proporsi varians dengan menggunakan rumus sebagai berikut:

$$\frac{\lambda_i}{\sum_{i=1}^{p} \lambda_i} \qquad (2.26)$$

- Menghitung proporsi kumulatif yang dapat dijelaskan oleh komponen utama ke- k dengan menggunakan rumus pada persamaan (2.27)

2.12. Pembentukan Komponen Utama

Syarat untuk membentuk komponen utama yang merupakan kombinasi linier dari variabel asal X agar mempunyai variansi maksimum adalah dengan memilih vektor eigen yaitu $\boldsymbol{v} = \left(\boldsymbol{v_1}, \boldsymbol{v_2}, \dots, \boldsymbol{v_p}\right)'$ sedemikian sehingga $var(K_i) = \boldsymbol{v_i}'\sum\boldsymbol{v_i}$ maksimum dengan batasan $\boldsymbol{v}'_{\boldsymbol{i}}\boldsymbol{v_i} = 1$.

- Komponen utama pertama adalah kombinasi linier $\boldsymbol{v}_1'\boldsymbol{X}$ yang memaksimumkan $var(\boldsymbol{v}_1'\boldsymbol{X})$ dengan syarat $\boldsymbol{v_1}'\boldsymbol{v_1} = \mathbf{1}$.
- Komponen utama kedua adalah kombinasi linier $\boldsymbol{v}_2'\boldsymbol{X}$ yang memaksimumkan $var(\boldsymbol{v}_2'\boldsymbol{X})$ dengan syarat $\boldsymbol{v_2}'\boldsymbol{v_2} = 1$ dan $cov(\boldsymbol{v}_1'\boldsymbol{X}, \boldsymbol{v}_2'\boldsymbol{X}) = 0$. Ini berarti komponen utama kedua tidak berkorelasi dengan komponen utama pertama.
- Komponen utama ke-k adalah kombinasi linier $\boldsymbol{v}_i'\boldsymbol{X}$ yang memaksimumkan $var\left(\boldsymbol{v}_j'\boldsymbol{X}\right)$ dengan syarat $\boldsymbol{v_j}'\boldsymbol{v_j} = 1$ dan $cov\left(\boldsymbol{v}_j'\boldsymbol{X}, \boldsymbol{v}_j'\boldsymbol{X}\right) = 0$ untuk $k < j$. Ini berarti komponen utama ke-k tidak berkorelasi dengan komponen utama sebelumnya.

Pembentukan komponen utama ke-$2,3, \dots, k$ dilakukan seperti pada langkah pembentukan komponen utama ke-1 sehingga diperoleh:

λ_j : nilai eigen ke-j terbesar dari Σ

$\boldsymbol{v_j}$: vektor eigen padanan λ_j

2.13. Kriteria Pemilihan Kompenen Utama

Salah satu tujuan dari *Analisis Komponen Utama* adalah mereduksi dimensi data asal yang semula terdapat p variabel bebas menjadi k komponen utama (dimana $k < p$). Kriteria pemilihan *k* (Johnson dan Wichern, 2007) yaitu:

1. Proporsi kumulatif keragaman data asal yang dijelaskan oleh k komponen utama minimal 80%.

$$\frac{\sum_{i=1}^{k} \lambda_j}{\sum_{i=1}^{p} \lambda_j} \tag{2.27}$$

2. Komponen utama yang dipilih adalah komponen utama yang mempunyai nilai eigen lebih besar satu atau $\lambda_i > 1$. Kriteria ini digunakan terutama untuk komponen utama yang diperoleh dari matriks korelasi. Alasan untuk membandingkannya dengan $\lambda_i > 1$ adalah ketika komponen utama yang diperoleh dari matriks korelasi (standarisasi data), variansi dari masing-

masing variabelnya sama dengan satu. Jika suatu komponen utama tidak dapat menerangkan variansi melebihi suatu variabel dapat menerangkan dirinya sendiri, maka komponen utama tersebut tidak signifikan atau dengan kata lain komponen utama yang mempunyai nilai eigen lebih kecil satu dapat diabaikan.

3. *Scree Plot*

   Scree Plot merupakan kriteria pemilihan jumlah komponen utama yang lebih bersifat grafis atau visual karena *Scree Plot* merupakan plot antara nilai eigen dan komponen utama yang terbentuk. Untuk menentukan jumlah komponen utama yang terbentuk perlu diperhatikan letak dimana terjadi patahan siku dari *Scree Plot* atau dipilih suatu titik terakhir dimana setelah titik tersebut plot cenderung meluruh (Rencher, 2002).

2.14. Teknik Kluster

Analisis cluster termasuk dalam analisis multivariat yang mewakili seluruh hubungan interdependensi, tidak ada perbedaan variabel bebas dan variabel tak bebas (*independent and dependent variables*) dalam analisis ini. “Analisis cluster adalah teknik yang digunakan untuk mengidentifikasi objek atau individu yang serupa dengan memperhatikan beberapa kriteria” (Kuncoro, 2003).

Analisis klaster adalah suatu teknik statistik yang bertujuan untuk mengelompokkan obyek ke dalam suatu kelompok sedemikian sehingga obyek yang berada dalam satu kelompok akan memiliki kesamaan yang tinggi dibandingkan dengan obyek yang berada di kelompok lain (Rahmawati, 2012). Dengan kata lain tujuan dari analisis cluster adalah pengklasifikasian obyek-obyek berdasar-kan similaritas diantaranya dan menghim-pun data menjadi beberapa kelompok (Anderberg, 2012). Terdapat dua macam metode dalam proses analisis *cluster* yaitu metode hirarki dan metode non hirarki.

Pada metode tidak berhirarki, digunakan jarak *Euclidian*, untuk menetapkan nilai kedekatan antara objek. Salah satu metode tidak berhirarki yaitu metode *K-means*. Metode *K-Means* pertama kali diperkenalkan oleh MecQueen JB pada tahun 1976. *K-Means* merupakan salah satu metode pengelompokan data tidak berhierarki (sekatan) yang beruaha mempartisi data yang ada kedalam bentuk dua atau lebih kelompok. Metode ini mempartisi data ke dalam kelompok sehingga data berkarakteristik sama di masukkan ke dalam satu kelompok yang sama dan data yang berkarakteristik berbeda dikelompokkan ke dalam kelompok yang lain (Sitepu, Irmeilyana, & Gulton, 2011).

Pada metode hirarki terdapat dua cara dalam analisis cluster yaitu dengan cara penggabungan*(agglomerative)* dan pemisahan *divisive*. Metode hirarki dengan cara penggabungan dilakukan dengan menggabungkan objek secara bertahap, sehingga nantinya hanya diperoleh satu kelompok saja. Sebaliknya cara pemisahan pada metode hirarki dimulai dengan membentuk satu kelompok/*cluster* besar yang anggotanya seluruh objek pengamatan. Kemudian dipisah menjadi kelompok yang lebih kecil, hingga akhirnya satu kelompok hanya beranggotakan satu objek pengamatan saja (Alwi & Hasrul, 2018)

Dalam penelitian ini menggunakan *Average Linkage. Average linkage* merupakan Jarak antara dua cluster diukur dengan jarak rataan antara sebuah objek dalam cluster yang satu dengan sebuah objek dalam cluster yang lain.

$$d_{(UV)W} = \frac{\sum_i \sum_d d_{ik}}{N_{(UV)} N_W} \tag{2.29}$$

Dimana d_{ik} jarak antara obyek *i* pada cluster (*UV*) dan obyek *k* pada *cluster* dalam klaster W, dan Nuv dan Nw berturut-turut adalah banyaknya item-item dalam klaster (UV) dan W (Johnson & Wichern, 2007).

2.15. Peubah *Dummy*

Peubah *dummy* atau peubah indikator adalah peubah buatan yang dibuat untuk mewakili atribut dengan dua kategori atau kategori yang berbeda. Variabel dummy menetapkan angka “0” dan “1” untuk menunjukkan keanggotaan dalam kategori yang saling ekskludif dan menyeluruh.Jumlah peubah *dummy* yang diperlukan untuk mewakili variabel atribut tunggal sama dengan jumlah kategori kurang satu. Untuk variabel atribut tertentu, tidak ada peubah *dummy* yang dibangun dapat berulang. Artinya, satu peubah *dummy* tidak bisa menjadi banyak konstanta atau hubungan linier sederhana yang lain. peubah dummy adalah peubah bebas yang mengambil nilai “0” atau “1”. Dalam model regresi, peubah *dummy* dengan nilai 0 akan menyebabkan koefisiennya hilang dari persamaan. Sebaliknya, nilai 1 menyebabkan koefisien berfungsi sebagai intercept tambahan, karena adanya properti identitas perkalian dengan 1. Jenis spesifikasi dalam model regresi linier ini berguna untuk menentukan himpunan bagian pengamatan yang memiliki kemiringan yang berbeda tanpa diciptakannya model terpisah (Parjiono *et al.*, 2018).