PERBANDINGAN METODE *BACKWARD ELIMINATION, FORWARD SELECTION* DAN *STEPWISE REGRESSION* DALAM PEMILIHAN MODEL REGRESI LINEAR BERGANDA TERBAIK

(Skripsi)

Oleh

YULICA DIAH KHOERUNNISA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG

2021

ABSTRACT

COMPARISON OF BACKWARD ELIMINATION, FORWARD SELECTION AND STEPWISE REGRESSION METHODS IN THE BEST MULTIPLE LINEAR REGRESSION MODEL SELECTION

By

YULICA DIAH KHOERUNNISA

The best regression model is a model that can explain the behavior of the dependent variable as well as possible, by selecting the independent variables from the many existing independent variables. One of the violations of assumptions in multiple regression analysis is the occurrence of a strong relationship between two independent variables called multicollinearity. The consequence of the multicollinearity problem is that the ordinary small squares estimate still meets the Best Linear Unbiased Estimator (BLUE) requirements, but the estimation becomes inefficient, therefore an improvement is needed to overcome this multicollinearity problem. The purpose of this study is to compare the methods of backward elimination, forward selection and stepwise regression in selecting the best multiple linear regression model in the case of multicollinearity using the coefficient of determination, $R^2_{Adjusted}$ and MSE . The results of this study indicate that the backward elimination, forward selection and stepwise regression methods can overcome multicollinearity and obtain the same linear regression model, so there is no difference in the best regression model in the cases studied in this study.

Keywords : Regression model, Multicollinearity, Backward Elimination, Forward Selection, Stepwise Regression

ABSTRAK

PERBANDINGAN METODE *BACKWARD ELIMINATION, FORWARD SELECTION* DAN *STEPWISE REGRESSION* DALAM PEMILIHAN MODEL REGRESI LINEAR BERGANDA TERBAIK

Oleh

YULICA DIAH KHOERUNNISA

Model regresi terbaik adalah model yang dapat menjelaskan perilaku variabel tak bebas dengan sebaik-baiknya, dengan memilih variabel-varibel bebas dari sekian banyak variabel bebas yang ada. Salah satu pelanggaran asumsi dalam analisis regresi berganda adalah terjadinya hubungan yang kuat antara dua variabel bebas yang disebut multikolinearitas. Konsekuensi dari masalah multikolinearitas yaitu penaksiran kuadrat kecil biasa masih tetap memenuhi syarat *Best Linear Unbiased Estimator* (BLUE), tetapi penaksiran tersebut menjadi tidak efisien, oleh sebab itu dibutuhkan sebuah perbaikan untuk mengatasi masalah multikolinearitas ini. Tujuan dari penelitian ini adalah membandingkan metode *backward eliminiation, forward selection dan stepwise regression* dalam pemilihan model regresi linear berganda terbaik dalam kasus multikolinearitas dengan menggunakan nilai koefisien determinasi, $R^2_{Adjusted}$ dan MSE . Hasil dari penelitian ini menunjukkan bahwa metode *backward eliminiation, forward selection dan stepwise regression* dapat mengatasi multikolinearitas dan memperoleh model regresi linear yang sama, sehingga tidak terdapat perbedaan model regresi terbaik pada kasus yang diteliti dalam penelitian ini.

Kata kunci : Model regresi, Multikolinearitas, *Backward Elimination, Forward Selection, Stepwise Regression*

PERBANDINGAN METODE *BACKWARD ELIMINATION, FORWARD SELECTION* DAN *STEPWISE REGRESSION* DALAM PEMILIHAN MODEL REGRESI LINEAR BERGANDA TERBAIK

Oleh

YULICA DIAH KHOERUNNISA

Skripsi

Sebagai Salah Satu Syarat untuk Mencapai Gelar
SARJANA MATEMATIKA

Pada

Jurusan Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Lampung

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2021

Judul Skripsi	:	**PERBANDINGAN METODE *BACKWARD ELIMINATION, FORWARD SELECTION* DAN *STEPWISE REGRESSION* DALAM PEMILIHAN MODEL REGRESI LINEAR BERGANDA TERBAIK**
Nama Mahasiswa	:	Yulica Diah Khoerunnisa
Nomor Pokok Mahasiswa	:	1717031011
Program Studi	:	Matematika
Fakultas	:	Matematika dan Ilmu Pengetahuan Alam

MENYETUJUI

1. Komisi Pembimbing

Drs. Eri Setiawan, M.Si.	**Dr. Muslim Ansori, S.Si., M.Si.**
NIP. 19581101 198803 1 002	NIP. 19720227 199802 1 001

2. Ketua Jurusan Matematika

Dr. Aang Nuryaman, S.Si., M.Si.
NIP. 19740316 200501 1 001

MENGESAHKAN

1. Tim Penguji

   Ketua : **Drs. Eri Setiawan, M.Si.** .....

   Sekretaris : **Dr. Muslim Ansori, S.Si., M.Si** .....

   Penguji
   Bukan Pembimbing : **Widiarti, S.Si., M.Si.** .....

2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam

Dr. Eng. Suripto Dwi Yuwono, S.Si., M.T.
NIP. 19740705 200003 1 001

Tanggal Lulus Ujian Skripsi : **09 Agustus 2021**

PERNYATAAN SKRIPSI MAHASISWA

Yang bertanda tangan dibawah ini:

Nama Mahasiswa	:	**Yulica Diah Khoerunnisa**
Nomor Pokok Mahasiswa	:	**1717031011**
Jurusan	:	**Matematika**
Judul Skripsi	:	**PERBANDINGAN *METODE BACKWARD ELIMINATION, FORWARD SELECTION* DAN *STEPWISE REGRESSION* DALAM PEMILIHAN MODEL REGRESI LINEAR BERGANDA TERBAIK**

Dengan ini menyatakan bahwa penelitian ini adalah hasil pekerjaan saya sendiri dan apabila kemudian hari terbukti bahwa skripsi ini merupakan hasil salinan atau dibuat oleh orang lain, maka saya bersedia menerima sanksi sesuai dengan ketentuan akademik yang berlaku.

Bandar Lampung, 09 Agustus 2021

Penulis



Yulica Diah Khoerunnisa

RIWAYAT HIDUP

Penulis bernama Yulica Diah Khoerunnisa dilahirkan di Gadingrejo pada tanggal 10 Juli 1999. Penulis merupakan anak keempat dari empat bersaudara dari pasangan Bapak Suyanto dan Ibu Sri Rahayu.

Penulis menempuh pendidikan di Sekolah Dasar (SD) Negeri 2 Tegalsari pada tahun 2005-2011, melanjutkan ke jenjang Sekolah Menengah Pertama (SMP) Negeri 1 Gadingrejo pada tahun 2011-2014, dan melanjutkan ke jenjang Sekolah Menengah Atas (SMA) Negeri 1 Gadingrejo pada tahun 2014-2017.

Pada tahun 2017 penulis terdaftar sebagai mahasiswa di jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung melalui jalur SNMPTN. Pada awal tahun 2020 di bulan Januari-Februari penulis melaksanakan Kuliah Kerja Nyata (KKN) di Desa Pampangan Kecamatan Sekincau Kabupaten Lampung Barat selama 40 hari. Pada bulan Juli-Agustus tahun 2020 penulis melakukan Kuliah Praktik (KP) di Badan Pusat Statistika (BPS) Kabupaten Pringsewu.

KATA INSPIRASI

“Barang siapa bertaqwa kepada Allah niscaya Dia akan membukakan jalan keluar baginya. Dan dia memberinya rezeki dari arah yang tidak disangka-sangkanya. Dan barang siapa bertawakal kepada Allah, niscaya Allah akan mencukupkan (keperluan)nya.”
(Q.S. At-Talaq:2-3)

“Allah tidak membebani seseorang melainkan sesuai dengan kesanggupannya.”
(Q.S. Al-Baqarah:286)

“Maka sesungguhnya bersama kesulitan ada kemudahan.”
(Q.S. Al-Insyirah:5)

PERSEMBAHAN

Dengan mengucap puji dan syukur kehadirat Allah SWT yang telah memberikan petunjuk dan kemudahan dalam menyelesaikan studiku dan kupersembahkan karyaku yang sederhana ini kepada

Alm. Bapak Suyanto dan Almh. Ibu Sri Rahayu

Bapak dan Ibu adalah sosok orangtua terbaik yang akan selalu menjadi kebanggaan anak-anaknya. Terimakasih sudah selalu memberikan yang terbaik untuk anak-anaknya.

Mb Ita, Mb Nurul, Mas Awan dan seluruh keluarga besar

Yang sudah selalu memberikan semangat, menjaga, menjadi pelindung, panutan serta menjadi pengganti sosok orangtua selama ini. Terimakasih atas waktu, kerja kerasnya, nasihatnya dan semangatnya untuk adikmu ini.

Guru dan Dosen

Terimakasih atas semua bimbingan dan ilmu pengetahuan yang diberikan kepadaku.

Almamater Tercinta Universitas Lampung

SANWACANA

Puji syukur kehadirat Allah SWT yang telah melimpahkan rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi dengan judul "**Perbandingan Metode *Backward Elimination, Forward Selection* dan *Stepwise Regression* dalam Pemilihan Model Regresi Linear Berganda Terbaik**". Penulis menyadari bahwa skripsi ini tidak akan terwujud tanpa adanya bantuan dan bimbingan dari berbagai pihak sehingga skripsi ini dapat terselesaikan. Pada kesempatan ini penulis mengucapakan terimakasih kepada :

1. Bapak Drs. Eri Setiawan, M.Si., selaku Pembimbing I yang telah memberikan arahan, bimbingan, dan saran kepada penulis sehingga dapat menyelesaikan skripsi ini.
2. Bapak Dr. Muslim Ansori, S.Si., M.Si., selaku Pembimbing II yang telah memberikan arahan dan saran dalam proses penyusunan skripsi ini.
3. Ibu Widiarti, S.Si., M.Si., selaku penguji yang telah memberikan kritik, saran dan evaluasi kepada penulis agar lebih baik lagi.
4. Ibu Dr. Asmiati, S.Si., M.Si., selaku pembimbing akademik yang telah memberikan bimbingan dan arahan selama perkuliahan ini.
5. Bapak Dr. Aang Nuryaman, S.Si., M.Si., selaku Ketua Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
6. Bapak Dr. Eng. Suripto Dwi Yuwono, S.Si., M.T., selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
7. Seluruh Dosen, staf, dan Karyawan Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung yang telah memberi ilmu dan bantuan kepada penulis.

8. Alm. Bapak, Almh Ibu, Mb Ita, Mb Nurul dan Mas Awan yang selalu memberikan do'a, semangat, nasihat dan motivasi untuk penulis.
9. Unni Zaa, Kakak Alleira, Aci Fiorin, Adek Amine, Kakak Bila, Abang Alzam, Anisya dan seluruh keluarga besar selalu mendoakan, mendukung dan memberikan semangat.
10. Nova, Zamhara, Joevita, dan Novela saudara beda ayah dan ibu yang selalu menemani disaat suka maupun duka dan selalu memberi semangat, dukungan dan motivasi selama ini.
11. Sahabat-sahabat Penulis Rafa, Chaterina, Stefani, Umroh, Dini, Nabilla, Beta dan Widiya yang telah membrikan banyak dukungan dan semangat.
12. Temen-temen seperbimbingan yang telah sama-sama berjuang dan memberikan semangat satu sama lain.
13. Teman-teman Mahasiswa Jurusan Matematika FMIPA Unila Angkatan 2017.
14. Serta seluruh pihak yang terkait yang telah banyak membantu yang tidak dapat disebutkan satu persatu.

Penulis menyadari bahwa skripsi ini jauh dari kata sempurna, sehingga penulis mengharapkan kritik dan sarannya yang dapat membangun kedepannya. Serta penulis juga berharap skripsi ini dapat bermanfaat untuk orang lain.

Bandar Lampung, 09 Agustus 2021
Penulis,

Yulica Diah Khoerunnisa

DAFTAR ISI

Halaman

DAFTAR TABEL xiv

I. PENDAHULUAN
1.1 Latar Belakang dan Masalah 1
1.2 Tujuan Penelitian 3
1.3 Manfaat Penelitian 3

II. TINJAUAN PUSTAKA
2.1 Analisis Regresi 4
2.2 Analisis Regresi Berganda 5
2.3 Asumsi Regresi Linear Berganda 6
2.4 Multikolinearitas 8
2.5 Konsekuensi Multikolinearitas 9
2.6 Metode Eliminasi Langkah Mundur (*Backward Elimination*) 12
2.7 Metode Seleksi Maju (*Forward Selection*) 12
2.8 Metode *Stepwise Regression* 13
2.9 Uji Kofisien Regresi Serentak (Uji F) 14
2.10 Uji Koefisien Regresi Secara Parsial (Uji t) 15
2.11 Pendeteksi Kelayakan Model Regresi Terbaik 16
2.11.1 Koefisien Determinasi 16
2.11.2 Nilai *Adjusted R-Square* 17
2.11.3 *Mean Square Error (MSE)* 18
2.12 Penduduk Miskin 18

III. METEDOLOGI PENELITIAN
3.1 Waktu dan Tempat Penelitian 20
3.2 Data Penelitian 20
3.3 Metode Penelitian 23

IV. HASIL DAN PEMBAHASAN
4.1 Hasil Simulasi untuk Kelompok Data dengan n=25, 50 dan 75 26
4.1.1 Nilai Korelasi dan VIF dengan n=25 n=25, 50 dan 75 26

4.1.2 Model Persamaan Regresi Linear untuk Data Simulasi 30
4.2 Hasil dari Data Real dengan Variabel Bebas Sebanyak 6 dan n=25 33
4.2.1 Nilai Korelasi dan VIF pada Data Presentase Penduduk Miskin di Provinsi Papua Tahun 2019..... 33
4.2.2 Model Regresi Linear dengan Metode *Backward Elimination* pada Data Persentase Penduduk Miskin Provinsi Papua Tahun 2019..... 36
4.2.3 Model Regresi Linear dengan Metode *Forward Selection* pada Data Persentase Penduduk Miskin Provinsi Papua Tahun 2019..... 41
4.2.4 Model Regresi Linear dengan Metode *Stepwise Regression* pada Data Persentase Penduduk Miskin Provinsi Papua Tahun 2019..... 43
4.2.5 Uji Asumsi Klasik pada Model Hasil Metode *Backward Elimination*, *Forward Selection* dan *Stepwise Regression*..... 47
4.2.5.1 Normalitas 48
4.2.5.1 Heteroskedastisitas 48
4.2.5.1 Multikolinearitas..... 49
4.2.6 Perbandingan Metode *Backward Elimination*, *Forward Selectio,* dan *Stepwise Regression*..... 50

V. KESIMPULAN..... 51

DAFTAR PUSTAKA 52

LAMPIRAN

DAFTAR TABEL

Tabel Halaman

1. Simulasi Data X_{ip} untuk 6 Variabel Bebas 21

2. Variabel Terikat (Y) untuk 6 Variabel Bebas 22

3. Data Persentase Penduduk Miskin Provinsi di Papua tahun 2019.....22

4. Korelasi antarvariabel Bebas dengan n=25.....26

5. Nilai VIF dengan n=25 27

6. Korelasi antarvariabel Bebas dengan n=50.....28

7. Nilai VIF dengan n=50 28

8. Korelasi antarvariabel Bebas dengan n=75.....29

9. Nilai VIF dengan n=75 30

10. Model Persamaan Regresi Linear 31

11. Model Persamaan Regresi Linear Metode *Backward Elimination, Forward Selection* dan *Stepwise Regression* untuk n=25.....31

12. Model Persamaan Regresi Linear Metode *Backward Elimination, Forward Selection* dan *Stepwise Regression* untuk n=50.....32

13. Model Persamaan Regresi Linear Metode *Backward Elimination, Forward Selection* dan *Stepwise Regression* untuk n=75.....32

14. Korelasi Variabel Terikat dan Variabel Bebas pada Data Persentase Penduduk Miskin di Provinsi Papua tahun 2019 34

15. Korelasi antarvariabel Bebas pada Data Persentase Penduduk Miskin Provinsi Papua tahun 2019.....35

16. Nilai VIF Data Persentase Penduduk Miskin Provinsi Papua Tahun 2019 35

17. Model *Summary* 37

18. Variabel Bebas yang dieliminasi dalam Model Regresi Metode *Backward Elimination*.....38

19. ANOVA untuk Metode *Backward Elimination*.....38

20. Uji Parsial untuk Metode *Backward Elimination* 40

21. Variabel Bebas yang dimasukkan dalam Model Regresi Metode *Forward Selection*.....42

22. ANOVA untuk Metode *Forward Selection* 42

23. Variabel Bebas yang dimasukkan/dieliminasi dalam Model Regresi Metode *Stepwise Regression*.....44

24. ANOVA untuk Metode *Stepwise Regression* 45

25. Uji Parsial untuk Metode *Stepwise Regression*.....46

26. Uji Normalitas.....48

27. Uji Heteroskedastisitas.....49

28. Uji Multikolinearitas 49

29. Perbandingan Metode *Backward Elimination, Forward Selection* dan *Stepwise Regression* 50

I. PENDAHULUAN

1.1 Latar Belakang dan Masalah

Analisis regresi adalah suatu analisis yang digunakan untuk menganalisis ketergantungan satu variabel dependen (terikat) dengan satu variabel atau lebih variabel independen (bebas), dengan tujuan untuk mengestimasi atau memprediksi nilai rata-rata variabel dependen berdasarkan nilai variabel independen yang diketahui. Apabila digunakan satu variabel independen maka regresi yang terbentuk disebut regresi sederhana. Apabila terdapat lebih lebih dari satu variabel independen maka regresi yang terbentuk disebut regresi berganda. Regresi linear adalah regresi dimana variabel independen dan dependennya memiliki pola hubungan yang linear.

Analisis regresi linear berganda merupakan analisis regresi linear yang melibatkan lebih dari satu variabel independen (Kutner et.al, 2005). Dalam menentukan model regresi, variabel bebas dapat masuk dalam model secara bersama-sama atau satu persatu. Jika variabel bebas masuk dalam model secara bersama-sama maka perhitungan akan ringkas, akan tetapi tidak akan terlihat apa yang terjadi dalam perhitungan tersebut karena setiap variabel bebas yang masuk memberikan pengaruh yang berbeda, tergantung pada urutan variabel bebas tersebut yang masuk dalam model. Namun tidak berarti semua variabel yang masuk dalam model regresi menjadikan model tersebut model yang terbaik (Sembiring, 1995). Model regresi terbaik adalah model yang dapat menjelaskan perilaku peubah tak

bebas dengan sebaik-baiknya dengan memilih peubah-peubah bebas dari sekian banyak peubah bebas yang tersedia dalam data. Sebuah model regresi berganda dikatakan baik atau cocok apabila memenuhi asumsi-asumsi klasik, yaitu galat menyebar normal dengan rata-rata nol, ragam dari galat bersifat homogen, galat tidak mengalami autokorelasi, dan tidak terjadi multikolinearitas antar variabel bebas.

Salah satu asumsi yang memungkinkan tidak terpenuhi adalah terjadinya multikolinearitas, yaitu hubungan yang kuat antara dua variabel bebas. Jika ada masalah multikolinearitas maka kesimpulan yang didapat dari hasil pengujian untuk model regresi maupun untuk masing-masing peubah yang ada dalam model, seringkali tidak tepat. Oleh sebab itu, masalah multikolinearitas harus dihindari. Untuk pemilihan model terbaik terdapat beberapa metode yang biasa digunakan yaitu metode *backward elimination,* metode *forward selection* dan metode *stepwise regression*. Metode *backward elimination* merupakan pemodelan dengan memasukkan semua variabel kemudian dikeluarkan satu persatu dengan melakukan pengujian terhadap parameter-parameternya dengan menggunakan partial F *test*. *Forward selection* merupakan pemodelan memasukkan variabel bebas yang memiliki korelasi yang paling erat dengan variabel tak bebasnya (variabel yang paling potensial untuk memiliki hubungan linear dengan variabel terikat). Kemudian secara bertahap memasukkan variabel bebas yang potensial berikutnya dan nanti akan berhenti sampai tidak ada lagi variabel bebas yang potensial. Metode *stepwise regression* melibatkan dua jenis proses yaitu *forward selection* dan *backward elimination.*

Berdasarkan pernyataan diatas penulis tertarik untuk melihat perbandingan metode *backward elimination,* metode *forward selection* dan metode *stepwise regression* dalam pemilihan model regresi linear terbaik pada kasus multikolinearitas dengan menggunakan nilai koefisien determinasi, $R^2_{Adjusted}$ dan *Mean Square Error* (MSE).

1.2 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah sebagai berikut:

1. Untuk mengetahui prosedur pemilihan model regresi linear berganda pada kasus multikolinearitas dengan metode *backward elimination, forward selection* dan *stepwise regression*.
2. Untuk mengetahui perbandingan antara metode *backward elimination, forward selection* dan *stepwise regression* dalam pemilihan model regresi linear berganda terbaik pada kasus multikolinearitas dengan menggunakan nilai koefisien determinasi, $R^2_{Adjusted}$ dan MSE.
3. Menentukan variabel yang mempengaruhi angka persentase penduduk miskin dengan metode *backward elimination, forward selection* dan *stepwise regression*.

1.3 Manfaat Penelitian

Adapun manfaat yang diharapkan dari penelitian ini adalah menambah pengetahuan bagi penulis dan memberikan masukan kepada peneliti dan pembaca dalam pemilihan model regresi linear berganda terbaik pada kasus multikolinearitas dengan menggunakan metode *backward elimination, forward selection* dan *stepwise regression*.

II. TINJAUAN PUSTAKA

2.1 Analisis Regresi

Analisis regresi merupakan suatu metode dalam statistika yang dapat digunakan untuk menyatakan hubungan fungsional antara dua variabel atau lebih. Dalam analisis regresi variabel dibedakan menjadi dua jenis yaitu variabel bebas dan juga variabel terikat. Varibel terikat adalah variabel yang akan diestimasi nilainya dan biasa diplot pada sumbu tegak (sumbu-Y). Sedangkan variabel bebas adalah variabel yang disumsikan memberikan pengaruh terhadap variansi variabel terikat dan biasa diplot pada sumbu datar (sumbu-X).

Misalkan diasumsikan model hubungan antara variabel X dan Y adalah linear dan ingin menentukan garis dugaan terbaiknya, maka harus menyadari bahwa garis dugaan dari masalah yang sebenarnya diharapkan mampu memprediksi dengan tepat setiap individu Y oleh setiap individu X. Aspek yang sangat penting dari analisis regresi adalah pengumpulan data karena kesimpulan dari analisis sangat tergantung pada data yang dikumpulkan. Pengumpulan data yang baik akan memberikan banyak manfaat, termasuk penyederhanaan analisis dan membangun

model yang secara umum dapat dipergunakan dan dipertanggungjawabkan (Usman dan Warsono, 2000).

Berdasarkan banyak variabel bebas yang digunakan, analisis regresi linear dibagi menjadi dua yaitu, analisis regresi linear sederhana dan analisis regresi linear berganda. Menurut Myers dan Milton (1991), analisis regresi yang hanya melibatkan satu peubah bebas disebut analisis regresi linear sederhana, sedangkan analisis regresi linear dengan peubah respon dipengaruhi oleh lebih dari satu peubah bebas disebut analisis regresi linear berganda.

2.2 Analisis Regresi Linear Berganda

Analisis Regresi linear berganda adalah salah satu metode statistika yang digunakan untuk mengetahui hubungan sebuah variabel terikat dengan beberapa variabel bebas. Tujuan analisis regresi linear berganda sama seperti dengan regresi linear sederhana, yaitu menggunakan hubungan antara variabel terikat dan variabel bebas untuk memprediksi atau menjelaskan karakterisristik dari variabel terikat (Freund, dkk, 2006). Persamaan umum garis regresi untuk regresi linear berganda adalah:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \varepsilon_i \qquad (2.1)$$

dengan,

- i = banyaknya pengamatan dengan $i = 1,2,3,\dots,n.$
- j = banyaknya variabel bebas dengan $j = 1,2,3,\dots,k.$
- Y_i = variabel terikat pengamatan ke-i.
- X_{ki} = variabel bebas pengamatan ke-i.
- β_0 = konstanta (parameter).
- β_k = koefisien regresi atau *slope* (parameter) ke-k.
- ε_i = sisaan (galat) pengamatan ke-i.

Dalam regresi linear berganda yang akan diduga adalah β_0 dan β_k artinya $(\beta_1, \beta_2, \dots, \beta_k)$. Persamaan linear untuk pendugaan garis regresi linear ditulis dalam bentuk:

$$\hat{y}_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + \cdots + b_k x_{ki} \qquad (2.2)$$

dengan,

$\hat{y}_i$ = nilai dugaan variabel terikat pengamatan ke-i.

x_{ki} = nilai variabel bebas pengamatan ke-i.

b_0 = titik potong regresi pada sumbu-y atau nilai dugaan $\hat{y}$ bila x=0.

b_k = gradient garis regresi (perubahan nilai dugaan $\hat{y}$ per satuan perubahan nilai x) ke-k.

Model regresi linear berganda dapat juga ditulis dalam bentuk matriks yaitu :

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \qquad (2.3)$$

dengan,

Y = vektor pengamatan berukuran $n \times 1$

X = matriks variabel bebas berukuran $n \times k$

β = vektor parameter yang akan ditaksir berukuran $k \times 1$

ε = vektor galat berukuran $n \times 1$

2.3 Asumsi Regresi Linear Berganda

Uji asumsi klasik terhadap model regresi linear berganda dilakukan agar dapat mengetahui bahwa model tersebut baik atau tidak untuk digunakan. Tujuan pengujian asumsi ini adalah untuk memberikan kepastian bahwa persamaan regresi yang diperoleh memiliki ketepatan dalam estimasi, tidak bias dan konsisten. Sehingga, sebelum melakukan analisis regresi berganda maka terlebih dahulu dilakukan uji asumsi.

Menurut Chatterjee dan Hadi (2006), untuk melakukan analisis regresi linear berganda harus memenuhi asumsi-asumsi berikut:

1. Normalitas

   Asumsi normalitas harus terpenuhi untuk mengetahui apakah residual atau galat dari data berdistribusi normal atau mengetahui apakah data sampel berasal dari populasi berdistribusi normal. Uji statistik yang sering digunakan adalah Kolmogorov-Smirnov.

2. Autokorelasi

   Autokorelasi terjadi karena galat antar pengamatan tidak saling bebas dan berkaitan antara satu dengan lainnya pada model regresi. Persyaratan yang harus dipenuhi adalah tidak adanya autokorelasi dalam model regresi.

3. Homoskedastisitas

   Homoskedastisitas dapat diartikan sebagai distribusi dari galat yang memiliki ragam konstan (homogen). Apabila varian galat dalam model tidak konstan disebut heteroskedastisitas. Heteroskedastisitas disebabkan karena variabel yang digunakan memiliki nilai yang sangat beragam, sehingga menghasilkaan nilai galat yang tidak konstan.

4. Multikolinearitas

   Salah satu syarat yang harus dipenuhi dalam pembentukan model regresi dengan beberapa variabel bebas adalah tidak adanya kasus multikolinearitas, yaitu tidak terdapat korelasi antara satu variabel bebas dengan variabel bebas lainnya. Dalam model regresi, adanya korelasi antar variabel bebas menyebabkan taksiran parameter regresi yang dihasilkan akan memiliki galat yang sangat besar. Pendeteksian kasus multikolinnearitas dilakukan menggunakan kriteria VIF (*Varians Inflation Factor*) lebih dari 10 menunjukkan adanya multikolinearitas antar variabel bebas.

2.4 Multikolinearitas

Istilah multikolinearitas pertama kali ditemukan oleh Ragnar Frisch Pada tahun 1034 yang artinya terdapat hubungan linear diantara beberapa atau semua variabel bebas pada model regresi. Masalah multikolinearitas hanya ditemukan dalam regresi linear berganda. Model yang baik adalah model yang bebas dari multikolinearitas. Suatu model yang bebas dari multikolinearitas adalah model yang memiliki nilai faktor *Variance Inflation Factor* (VIF) < 10. Nilai VIF > 10 mengindikasi terdapat multikolinearitas (Myers, 1990).

Salah satu cara untuk menguji gejala multikolinearitas dalam model regresi adalah dengan melihat nilai TOL (*tolerance*) dan VIF (*Variance Inflation Factor*) masing-masing variabel bebas terhadap variabel terikatnya (Gujarati, 1995). Uji untuk mengetahui gejala multikolinearitas dengan melihat nilai VIF dan TOL tersebut dilakukan dengan menggunkan langkah-langkah:

1. Menghitung VIF dari X_1.
2. Meregresikan variabel bebas selain X_1 terhadap X_1.
3. Menghitung koefisien determinasi dari regresi variabel bebas selai X_1 terhadap X_1 dan diperoleh R_j^2.
4. Menghitung nilai TOL dengan rumus $TOL = (1 - R_j^2)$.
5. Mengitung nilai VIF dengan rumus $VIF = \frac{1}{Tol}$.

Masalah multikolinearitas bisa timbul karena berbagai sebab. Pertama, karena sifat-sifat yang terkandung dalam kebanyakan variabel ekonomi berubah bersama-sama sepanjang waktu. Besaran-besaran ekonomi dipengaruhi oleh faktor faktor yang sama. Oleh karena itu, sekali faktor-faktor yang mempengaruhi menjadi operatif, maka seluruh variabel akan cenderung berubah dalam satu arah. Dalam data *time series*, pertumbuhan dan faktor-faktor kecenderungan merupakan penyebab utama adanya multikolinearitas. Kedua penggunaan nilai lag (*lagget*

values) dari variabel-variabel bebas tertentu dalam model regresi. Mengingat sifat yang sangat mendasar dari data, multikolinearitas diperkirakan terdapat pada sebagian besar hubungan-hubungan ekonomi. Oleh karena itu, perhatian sesungguhnya bukan lagi terletak pada ada atau tidaknya multikolinearitas, tetapi lebih pada akibat-akibat yang ditimbulkan oleh adanya multikolinearitas dalam sampel (Sumodiningrat, 1998).

2.5 Konsekuensi Multikolinearitas

Jika asumsi pada model linear klasik terpenuhi, maka penaksir kuadrat terkecil atau *Ordinary Least Square* dari koefisien regresi linear adalah linear, tak bias dan mempunyai varian minimum dalam arti penaksiran tersebut adalah penaksiran tak bias kolinear terbaik atau *Best Linear Unbiased Estimator* (BLUE), meskipun multikolinearitas sangat tinggi, penaksiran kuadrat terkecil biasa masih tetap memenuhi syarat BLUE, tetapi penaksiran tersebut tidak stabil (Gujarati, 1995).

Multikolineaaritas berpengaruh terhadap estimasi kuadrat terkecil dari koefisien regresi. Akan diperlihatkan bagaimana $\hat{\beta}$, variansi $(\hat{\beta}_j)$ dan kovariansi $(\hat{\beta}_j, \hat{\beta}_h)$ dengan $j, h = 1,2, \dots, k$. Misalkan ada dua variabel bebas (X_1, X_2) dan Y variabel terikat sehingga diperoleh model,

$$Y = \beta_1 X_1 + \beta_2 X_2 + \varepsilon_i \qquad (2.4)$$

Persamaan matriks dengan metode kuadrat terkecil adalah

$$(\boldsymbol{X^T X})\widehat{\boldsymbol{\beta}} = \boldsymbol{X^T Y} \qquad (2.5)$$

$$\boldsymbol{X^T X} = \begin{bmatrix} 1 & r_{12} \\ r_{21} & 1 \end{bmatrix}, \qquad \boldsymbol{X^T Y} = \begin{bmatrix} r_{1y} \\ r_{2y} \end{bmatrix}$$

diperoleh

$$[\boldsymbol{X^T X}]^{-1} = \begin{bmatrix} \frac{1}{1-r_{12}^2} & \frac{-r_{12}}{1-r_{12}^2} \\ \frac{-r_{21}}{1-r_{12}^2} & \frac{1}{1-r_{12}^2} \end{bmatrix}$$

Elemen diagonal utama dari matriks $[\boldsymbol{X^T X}]^{-1}$merupakan *Varians Inflasion Factor* (VIF) yaitu:

$$C_{jj} = \frac{1}{1-R_j^2} \; ; j = 1,2, \dots, k \qquad (2.6)$$

dengan,

C_{jj} = Varians Inflasion Factor (VIF).

R_j^2 = koefisien determinasi dari regresi X_j.

r_{12} = korelasi antara X_1 dan X_2.

r_{XjY} = korelasi antara X_j dan Y.

$$(\boldsymbol{X^T X})\widehat{\boldsymbol{\beta}} = \boldsymbol{X^T Y}$$

$$\begin{bmatrix} 1 & r_{12} \\ r_{21} & 1 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} r_{1y} \\ r_{2y} \end{bmatrix}$$

sehingga,

$$\hat{\beta}_1 = \frac{r_{1y} - r_{21} r_{2y}}{1 - r_{12}^2}$$

$$\hat{\beta}_2 = \frac{r_{2y} - r_{12} r_{1y}}{1 - r_{12}^2}$$

Jika ada multikolinearitas antara X_1 dan X_2 yang sangat erat dan $r_{12} \to 1$. Variansi dan kovariansi koefisien regresi menjadi sangat besar karena $V(\hat{\beta}_j) = C_{jjr2} \to \infty$ seperti $|r_{12}| \to 1$, galat $Cov(\hat{\beta}_1, \hat{\beta}_2) = C_{12}\delta^2 \to \pm\infty$, variansi yang besar untuk $\hat{\beta}_j$ menyatakan bahwa koefisien regresi adalah perkiraan yang sangat lemah. Pengaruh multikolinearitas adalah untuk memperkenalkan sebuah ketergantungan linier yang dekat dalam kolom matriks.

Selanjutnya jika kita mengasumsikan $X_1^T Y \rightarrow X_2^T Y$, seperti $|r_{12}| \rightarrow 1$, perkiraan koefisien regresi menjadi sama besarnya, tetapi berlawanan tanda, yaitu $\hat{\beta}_1 = -\hat{\beta}_2$. Masalah yang sama terjadi bila masalah multikolinearitas disajikan pada lebih dari dua variabel bebas. Umumnya elemen diagonal matriks $C = [\boldsymbol{X^T X}]^{-1}$ dapat ditulis sebagai berikut:

$$C_{jj} = \frac{1}{1 - R_j^2}$$

R_j^2 dihasilkan dari meregresikan X_1 pada variabel bebas lainnya. Nilai VIF dapat dihitung dengan menggunakan cara:

$$VIF\ (\hat{\beta}_j) = \frac{1}{1 - R_j^2} \qquad (2.7)$$

Varians Inflasion Factor (VIF) untuk $\hat{\beta}_j$ini adalah ukuran penting dalam perkiraan multikolinearitas. Menurut Sumodiningrat (1998), dalam hal terdapat multikolinearitas sempurna, penaksir dengan kuadrat terkecil bisa menjadi tak tentu dan variansi serta standar deviasinya menjadi tak terhingga. Sedangkan jika multikolinearitas tinggi, tetapi tidak sempurna maka konsekuensinya sebagai berikut:

a. Meskipun penaksir melalui kuadrat terkecil bisa didapatkan, standar deviasinya cenderung lebih besar jika derajat kolinearitas antara peubah bertambah.
b. Karena standar deviasi besar, interval kepercayaan bagi parameter populasi yang relevan akan menjadi besar.
c. Taksira-taksiran parameter kuadrat terkecil didapatkan tetapi standar deviasi akan menjadi sangat sensitif terhadap perubahan.
d. Jika multikolinearitas tinggi, mungkin bisa tinggi namun tidak satu pun (sangat sedikit) taksiran koefisien regresi yang signifikan secara statistik.

2.6 Metode Eliminasi Langkah Mundur (*Backward Elimination*)

Metode *backward elimination merupakan* eliminasi langkah mundur adalah kebalikan dari metode seleksi maju, dimana pertama-tama kita memasukkan semua variabel bebas kedalam model lalu kita eliminasi satu persatu yang tidak sesuai. Pengeliminasian variabel bebas didasarkan pada nilai F_{hitung} dan turut tidaknya variabel bebas pada model juga ditentukan dengan nilai F_{tabel} .

2.7 Metode Seleksi Maju (*Forward Selection*)

Metode seleksi maju bekerja dengan cara memasukkan peubah bebas satu demi satu menurut urutan besar pengaruhnya terhadap model, kemudian menguji signifikansi parameter regresi dan proses berhenti bila semua peubah bebas yang memenuhi syarat telah masuk ke dalam model yang dibangun (Sembiring, 1995).

Pengaruh (korelasi) positif atau negatif tidak dipersoalkan karena yang diperhatikan hanyalah eratnya hubungan antara suatu peubah bebas dengan Y sedangkan arah hubungan tidak menjadi persoalan. Salah satu keuntungannya ialah dapat melihat proses pembentukkan model itu tahap demi tahap dimulai dari yang pertama kali (Sembiring, 1995).

Bila nilai F_{hitung} lebih kecil dari yang ditetapkan untuk pemasukkan peubah bebas ke dalam model maka X_j tidak jadi masuk, begitu juga sebaliknya.

2.8 Metode *Stepwise Regression*

Metode *stepwise regression* merupakan sebuah metode yang dapat digunakan dalam menentukan model regresi linear berganda. Metode *stepwise Regression* melibatkan dua jenis proses yaitu *forward selection* dan *backward elimination*. Prosedur yang dilakukan yaitu dengan menyusupkan satu demi satu variabel bebas sampai diperoleh persamaan regresi yang dibutuhkan. Urutan penyisipan ditentukan dari koefisien korelasi parsial (r) (Draper & Smith, 1992). Jadi, variabel bebas yang mempunyai korelasi terbesar dengan variabel terikat akan masuk pertama kali dalam model.

Apabila salah satu peubah telah dimasukkan ke dalam model regresi, maka peubah lainnya tidak perlu dimasukkan lagi ke dalam model regresi, karena pengaruhnya telah diwakili oleh peubah yang sudah masuk ke dalam model regresi. Sehingga tidak terdapat multikolinearitas pada model regresi yang dihasilkan (Sembiring, 1995).

Tahap penyelesaian pada metode *stepwise regression* yaitu membentuk matriks koefisien korelasi. Dimana koefisien regresi yang dibutuhkan adalah koefisien korelasi linear sederhana *Y* dan X_i dengan rumus:

$$r_{yx_i} = \frac{\sum(X_{ij} - \bar{X}_i)(Y_j - \bar{Y})}{\sqrt{\sum(X_{ij} - \bar{X}_i)^2 \sum(Y_j - \bar{Y})^2}} \qquad (2.8)$$

Oleh karena itu, dapat dibentuk matriks koefisien korelasi linier sederhana antara *Y* dan X_i.

$$\begin{bmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{21} & 1 & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & 1 \end{bmatrix}$$

Kemudian membentuk persamaan regresi linear yang pertama. Variabel yang pertama diregresikan adalah variabel yang mempunyai harga mutlak koefisien korelasi yang terbesar antara Y dan X_i, misalnya X_i.

$$X = \begin{pmatrix} 1 & X_{h1} \\ 1 & X_{h2} \\ \vdots & \vdots \\ 1 & X_{hn} \end{pmatrix} (X^T X)^{-1} = \begin{pmatrix} n & \sum X_h \\ \sum X_h & \sum X_h^2 \end{pmatrix}$$

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} X^T Y = \begin{pmatrix} \sum Y \\ \sum Y X_i \end{pmatrix}$$

Langkah selanjutnya yaitu seleksi variabel kedua diregresikan. Cara menyeleksi variabel yang diregresikan adalah memilih parsial korelasi variabel sisa terbesar. Untuk menghitung harga masing-masing parsial korelasi sisa digunakan rumus:

$$r_{YX_h.X_k} = \frac{r_{YX_h} - r_{YX_k} r_{X_{h.X_k}}}{\sqrt{\left(1 - r_{YX_k}^2\right)\left(1 - r_{X_{hX_k}}^2\right)}} \qquad (2.9)$$

Tahap selanjutnya adalah memasukkan variabel bebas yang memiliki korelasi terbesar kedua dengan variabel terikat. Tahap ini terus dilakukan sampai semua variabel masuk dalam model, kemudian membuat persamaan regresi kedua :

$$Y = b_0 + b_1 X_1 + \cdots + b_k X_k \qquad (2.10)$$

Setelah diperoleh satu variabel bebas yang masuk dalam model perlu dilakukan pengujian keberartian variabel dalam model.

2.9 Uji Kofisien Regresi Serentak (Uji F)

Uji F dilakukan untuk mengetahui pengaruh semua variabel bebasnya terhadap variabel terikat. Selain itu uji F juga dapat dilakukan untuk mengetahui

signifikansi koefisien determinasi. Persamaan uji statistik F adalah sebagai berikut:

$$F_{hitung} = \frac{\frac{R^2}{k-1}}{1-\frac{R^2}{n-k}} \quad (2.11)$$

dengan,

R^2 = koefisien determinasi

n = jumlah pengamatan

k = jumlah parameter yang diestimasi

Sedangkan hipotesis dalam uji F adalah:

$$H_0: \beta_1 = \beta_2 = \cdots = \beta_k = 0$$

$$H_1: \beta_1 \neq \beta_2 \neq \cdots \neq \beta_k \neq 0$$

Jika F_{hitung} lebih besar dari F_{tabel}, maka H_0 ditolak, artinya variabel bebas mempunyai pengaruh secara keseluruhan terhadap variabel terikat, demikian sebaliknya.

2.10 Uji Koefisien Regresi Secara Parsial (Uji t)

Uji ini dilakukan untuk mengetahui signifikansi variabel bebas secara individu terhadap variabel terikatnya. Hipotesis yang digunakan adalah sebagai berikut:

$$H_0: \beta_1 = 0$$

$$H_1: \beta_1 \neq 0$$

Kriteria uji yang digunakan adalah jika $|t_{hitung}| > t_{tabel}$ maka tolak H_0. Jika tolak H_0 maka terdapat pengaruh secara individu variabel bebas terhadap variabel terikatnya.

2.11 Pendeteksi Kelayakan Model Regresi Terbaik

Salah satu tujuan di dalam analisis regresi adalah untuk mendapatkan model terbaik yang menjelaskan hubungan antara variabel terikat dengan variabel bebas. Model terbaik adalah model yang seluruh koefisien regresinya berarti (signifikan) dan mempunyai kriteria model terbaik optimum. Suatu model regresi dikatakan layak dan terbaik apabila model tersebut memenuhi kriteria sebagai berikut:

2.11.1 Koefisien Determinasi

Koefisien determinasi (R^2) merupakan proporsi dari Jumlah Kuadrat Total (terkoreksi) dari variabel tidak bebas terhadap Jumlah Kuadrat Total dari variabel-variabel bebas dalam model:

$$R^2 = \left(\frac{n \sum XY - \sum X \sum Y}{(n \sum X^2 - (\sum X)^2)\,(n \sum Y^2 - (\sum Y)^2)} \right)^2 \qquad (2.12)$$

dengan,

n = Banyaknya pengamatan.
Y = variabel terikat.
X = variabel bebas.

Tujuannya adalah memilih suatu model dengan memperhatikan kenaikan R^2 yang menghasilkan perbedaan terbesar pada banyaknya variabel yang diambil. Apabila R^2 tidak mengalami penurunan pada setiap variabel bebas yang ditambahkan pada model, maka model telah memberikan R^2 maksimum yang telah cukup untuk model dari semua peubah bebas yang dikandungnya. Plot khusus dari R^2 terhadap jumlah variabel dalam model, akan menunjukkan kenaikan bertahap R^2 dari

dengan tingkat yang mendekati R^2 maksimum dengan variabel-variabel penting yang termasuk dalam model.

2.11.2 Nilai *Adjusted R-Square*

Karena adanya kelemahan dalam perhitungan R^2, banyak peneliti yang menyarankan untuk menggunakan *adjusted r-square*. Interpretasinya sama dengan *R-Square*, akan tetapi nilai *adjusted r-square* dapat naik atau turun dengan adanya penambahan variabel baru, tergantung dari korelasi antara variabel bebas tambahan tersebut dengan variabel terikatnya. Nilai *adjusted r-square* dapat bernilai negatif, sehingga jika nilainya negatif maka nilai tersebut dianggap nol, atau variabel bebas sama sekali tidak mampu menjelaskan varians dari variabel terikatnya.

Suatu sifat penting R^2 adalah nilainya merupakan fungsi yang tidak pernah menurun dari banyaknya variabel bebas yang ada dalam model. Oleh karena itu, untuk membandingkan dua R^2 dari dua model harus memperhitungkan banyaknya variabel bebas yang ada dalam model tersebut. Ini dapat dilakukan dengan menggunakan "*adjusted R-square*". Istilah penyesuaian berarti nilai R^2 sudah disesuaikan dengan banyaknya variabel (derajat bebas) dalam model. Memang, R^2 yang disesuaikan ini juga akan meningkat bersamaan dengan meningkatnya jumlah variabel, tetapi peningkatan relative kecil. Seringkali juga disarankan, jika variabel bebas lebih dari dua, sebaiknya menggunakan *adjusted r square*.

Menurut Sembiring (1995), statistik R^2 yang sering disebut dengan koefisien determinasi. Sedangkan *R^2 adjusted* memperhitungkan banyaknya variabel bebas dalam model serta telah disesuaikan terhadap derajat bebas masing-masing jumlah

kuadrat. Model yang baik memiliki R^2 *adjusted* yang besar. R^2 *adjusted* dirumuskan dengan:

$$R^2_{adjusted} = 1 - (1 - R^2)\frac{(n-1)}{(n-p)} \tag{2.13}$$

dengan,

R^2 = sampel R-square.

p = banyaknya parameter termasuk Y.

n = banyaknya pengamatan.

2.11.3 *Mean Square Error (MSE)*

Menurut Montgomery, Peck & Vining (2001), jika ada dua penaksir parameter yaitu penaksir bias dengan variansi yang kecil dan penaksir tak bias dengan variansi yang besar makan penaksir bias dengan variansi kecil lebih dipilih sebagai penaksir parameter. Ukuran yang digunakan untuk membandingkan kedua hasil taksiran adalah di atas disebut *mean square error* (MSE) . Misalkan $\hat{\beta}$ adalah penaksiran parameter β. MSE dari $\hat{\beta}$ yaitu didefinisikan oleh:

$$MSE\left(\hat{\beta}\right) = \frac{1}{m}\sum_{i=1}^{m}(\hat{\beta} - \beta)^2 \; ; j = 1,2, \dots, p \tag{2.14}$$

dengan m adalah banyaknya pengulangan dalam simulasi. Semakin kecil nilai MSE menunjukkan bahwa suatu metode semakin baik.

2.12 Penduduk Miskin

Secara umum, kemiskinan merupakan kondisi dimana seseorang atau sekelompok orang yang tidak mampu memenuhi hak-hak dasarnya untuk mempertahankan dan mengembangkan kehidupan yang bermartabat. Penduduk miskin adalah

penduduk yang memiliki rata-rata pengeluaran per kapita perbulan dibawah garis kemiskinan , yang diperoleh dari hasil survei.

Menurut BPS Provinsi Papua angka presentase penduduk miskin pada tahun 2019 apabila dibandingkan dengan tahun sebelumnya, memiliki nilai 27,74% pada tahun 2018 dan menjadi 27,53% pada tahun 2019. Pada tahun 2019 terdapat 2 kabupaten/kota yang memiliki nilai presentase penduduk miskin yang tinggi yaitu Deiyai (43,65%) dan Intan Jaya (42,92%). Kabupaten/kota dengan angka presentase tinggi masih perlu dilakukan upaya mengatasi kemiskinan.

III. METODOLOGI PENELITIAN

3.1 Waktu dan Tempat Penelitian

Penelitian ini dilakukan pada semester ganjil tahun akademik 2020/2021. Bertempat di Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung.

3.2 Data Penelitian

Data yang digunakan dalam penelitian ini adalah data simulasi dan data real yang mengandung multikolinearitas. Untuk data simulasi yakni data yang dibangkitkan dengan variabel bebas sebanyak p=6 dan banyaknya data yaitu n=25,50 dan 75.

Untuk mendapatkan data multikolinearitas pada setiap himpunan data X_p dibangkitkan menggunakan simulasi Monte Carlo berdasarkan McDonald dan Galarneau (1975) dengan persamaan sebagai berikut:

$$X_p = \sqrt{(1-\rho^2)z_{ij}} + \rho z_{i(p+1)} \qquad i = 1,2,\dots,n$$

$$j = 1,2,\dots p$$

dengan $z_{i1}, z_{i2}, \dots, z_{i(p+1)}$ merupakan data yang dibangkitan dalam bentuk distribusi normal N(0,1) dan ρ ditentukan sehingga korelasi antara dua variabel penjelas diberikan oleh ρ^2. Dua himpunan dari variabel yang saling berkorelasi dalam penelitian ini dibuat berdasarkan nilai ρ_1 = 0.3 dan ρ_2 = 0.99. Untuk lebih jelasnya dapat dilihat pada tabel berikut.

Tabel 1. Simulasi Data X_{ip} untuk 6 variabel bebas

Jumlah Variabel Bebas	Simulasi Data X_{ip}
6	$Z_{ij} = N(0,1) \qquad j = 1, 2, \dots, 7$ $X_{i1} = (1-\rho_1^2)^{\frac{1}{2}} Z_{i1} + \rho_1 z_{i7}$ $X_{i2} = (1-\rho_1^2)^{\frac{1}{2}} Z_{i2} + \rho_1 z_{i7}$ $X_{i3} = (1-\rho_2^2)^{\frac{1}{2}} Z_{i3} + \rho_2 z_{i7}$ $X_{i4} = (1-\rho_1^2)^{\frac{1}{2}} Z_{i4} + \rho_1 z_{i7}$ $X_{i5} = (1-\rho_1^2)^{\frac{1}{2}} Z_{i5} + \rho_1 z_{i7}$ $X_{i6} = (1-\rho_2^2)^{\frac{1}{2}} Z_{i6} + \rho_2 z_{i7}$

Variabel terikat (Y) untuk setiap p variabel bebas diperoleh berdasarkan model $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$ dimana β adalah $\beta_{i,j} = 1$; untuk i=j, dan 0 selainnya. Dengan ε dibangkitkan berdasarkan distribusi normal N(0,1) sehingga Y merupakan kombinasi linear dari p variabel bebas ditambah galat yang ditampilkan pada tabel berikut.

Tabel 2. Variabel Terikat (Y) untuk 6 Variabel Bebas

p	Y
6	$Y = X_1 + X_2 + \cdots + X_p + N(0,1) \qquad p = 1,2,\ldots,6$

Sedangkan untuk data real menggunakan data sekunder yang berasal dari Badan Pusat Statistika (BPS) Provinsi Papua. Data yang digunakan adalah data Persentase Penduduk Miskin (PPM) di Provinsi Papua. Menurut BPS Provinsi Papua rata-rata persentase peduduk miskin berdasarkan kabupaten/kota tahun 2019 di Provinsi Papua sebesar 27.53%. Faktor-faktor yang mempengaruhi PPM adalah sebagai berikut:

X1= garis kemiskinan

X2= pendapatan per kapita

X3= rata-rata lama sekolah

X4= persentase penduduk bekerja

X5= jumlah penduduk miskin

X6= indeks pembangunan manusia

Tabel 3. Data Persentase Penduduk Miskin Provinsi di Papua tahun 2019

Kabupaten/Kota	Y	X1	X2	X3	X4	X5	X6
Merauke	10,35	346502	10498	8,56	97,39	23,49	69,98
Jayawijaya	38,33	420471	7835	5,3	97,61	83,23	57,79
Jayapura	13,13	589521	10375	9,79	90,32	17,19	71,84
Nabire	24,81	579753	9195	9,7	93,69	37,14	68,53
Kep. Yapen	27,13	618487	7785	9,19	94,22	27,19	67,76
Biak Numfor	25,5	543971	10211	10,22	89,58	38,6	72,57
Paniai	37,16	479877	6767	4,38	99,34	65,54	56,58
Puncak Jaya	35,71	604945	5523	3,61	98,22	45,89	48,33
Mimika	14,54	836301	12035	9,91	92,49	31,79	74,13
Boven Digoel	19,66	458012	8300	8,55	96,92	13,54	61,51
Mappi	25,5	330427	6513	6,3	95,49	26,1	58,3
Asmat	26,6	364728	6066	4,82	98,94	25,8	50,37
Yahukimo	38,82	414936	5030	4,02	97,99	73,92	49,25

Tabel 3. Lanjutan

Peg. Bintang	30,51	560354	5633	2,61	98,37	23,01	45,21
Tolikara	32,9	382936	5142	3,63	98,7	45,65	49,68
Keerom	16,83	655056	9136	8	97,1	9,55	66,59
Waropen	30,95	655060	7018	9,18	96,68	9,68	65,34
Mamberamo Raya	29,13	683779	4807	5,65	97	6,96	52,2
Nduga	38,24	341566	4181	0,97	98,97	37,6	30,75
Lanny Jaya	39,52	464328	4569	3,19	99,88	70,6	48
Mamberamo Tengah	36,93	388619	4671	2,9	99,29	17,79	47,23
Yalimo	34,52	330040	4860	2,58	99,36	21,48	48,08
Puncak	38,24	630074	5702	1,96	100	43,09	42,7
Dogiyai	31,12	492905	5709	4,92	99,89	30,37	55,41
Kota Jayapura	11,49	992610	15176	11,55	87,63	34,42	80,16

3.3 Metode Penelitian

Penelitian ini dilakukan secara studi pustaka yaitu mempelajari buku-buku penujang dan karya ilmiah yang disajikan dalam bentuk jurnal. Untuk mempermudah perhitungan dan mendapatkan hasil yang akurat menggunakan *software* R-Studio. Adapun langkah-langkah yang dilakukan dalam penelitian ini yaitu:

1. Menginput data real persentase penduduk miskin di Indonesia tahun 2019 dan melakukan simulasi data dengan membangkitkan data yang mengandung multikolinearitas dengan simulasi monte carlo dengan persamaan $X_p = \sqrt{(1-\rho^2)z_{ij}} + \rho z_{i(p+1)}$ dimana $i = 1,2,\dots,n$, $j = 1,2,\dots p$ dengan banyak variabel bebas (p) 6 dan banyaknya pengamatan (n) 25, 50 dan 75 dan $\rho_1 = 0.3$ dan $\rho_2 = 0.99$.
2. Melakukan standarisasi data dan membuat matriks korelasi.
3. Mengidentifikasi multikolinearitas dengan melihat nilai korelasi dan nilai VIF (*Variance Inflation Factor*) pada data simulasi dan data real.
4. Membentuk persamaan regresi linear metode *backward elimination* dengan langkah-langkah sebagai berikut:

a. Meregresikan seluruh variabel bebas dan variabel terikat, selanjutnya lihat pada uji parsial dan bandingkan signifikansi dan nilai t_{hitung} setiap variabel bebas.
b. Keluarkan variabel bebas yang memiliki nilai signifikansi besar dari α(0.05) dan nilai t_{hitung} lebih kecil dari t_{tabel} dari dari pesamaan regresi.
c. Meregresikan variabel bebas yang tersisa dengan variabel terikat.
d. Perhatikan hasil uji parsialnya, jika terdapat nilai nilai signifikansi besar dari α(0.05) dan nilai t_{hitung} lebih kecil t_{tabel} maka kembali kelangkah b . Sedangkan jika nilai signifikansi setiap variabel bebas lebih kecil dari α(0.05) dan nilai t_{hitung} lebih besar dari t_{tabel} maka lanjutkan ke langkah berikutnya.
e. Menentukan model yang signifikan secara parsial dan serentak.

5. Membentuk persamaan regresi linear dengan metode *forward selection* dengan langkah-langkah sebagai berikut:
   a. Menentukan nilai korelasi pearson setiap variabel bebas dengan variabel terikat.
   b. Membentuk regresi pertama dengan memasukkan variabel bebas yang mempunyai nilai korelasi terbesar antara variabel terikat dan variabel bebas.
   c. Menentukan korelasi parsial setiap variabel bebas yang belum masuk dalam model dengan variabel terikat.
   d. Membentuk regresi kedua dengan memasukkan variabel bebas yang mempunyai nilai korelasi parsial terbesar.
   e. Melakukan uji persamaan regresi linear apakah koefisien regresi tersebut signifikan atau tidak. Apabila nilai signifikansi lebih besar dari α(0.05), maka variabel bebas tersebut tidak masuk dalam model dan jika memberikan nilai signifikansi lebih kecil dari α(0.05), maka peubah bebas tersebut masuk ke dalam model.
   f. Jika semua variabel bebas belum masuk kedalam model maka ulangi langkah c sampai langkah e.
6. Membentuk persamaan regresi linear metode *stepwise regression* dengan langkah-langkah sebagai berikut:

a. Pemilihan variabel bebas pertama yang masuk kedalam model adalah variabel bebas yang memiliki nilai koefisien korelasi terbesar terhadap nilai variabel terikat.
b. Regresikan variabel terikat dan variabel bebas yang terpilih pada langkah a. dan melakukan uji-F dan uji-t dan nilai signifikansi $<\alpha(0.05)$.
c. Menentukan korelasi parsial setiap variabel bebas yang belum masuk dalam model dengan variabel terikat.
d. Membentuk regresi kedua dan melakukan uji keberartian regresi terhadap variabel. Variabel yang akan tetap ada dalam model adalah variabel bebas yang memiliki nilai signifikasi$< \alpha(0.05)$.
e. Jika semua variabel bebas belum masuk kedalam model maka ulangi langkah c sampai langkah d.

7. Melakukan uji asumsi model persamaan linear *backward elimination, forward selection dan stepwise regression.*
8. Membandingkan model persamaan linear metode *backward elimination, forward selection dan stepwise regression* dengan menggunakan nilai koefisien determinasi, *Adjusted R-Square,* dan MSE.
9. Membuat Kesimpulan.

V. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan dapat disimpulkan bahwa:

1. Metode *backward elimination, forward selection* dan *stepwise regression* dapat mengatasi multikolinearitas pada model regresi linear berganda yang mengalami kasus multikolinearitas.
2. Metode *backward elimination, forward selection* dan *stepwise regression* menghasilkan model regresi linear yang sama dalam mengatasi kasus multikolinearitas yaitu.

$$Y = 2.829669 \times 10^{-16} + 2.124118 \times 10^{-1}X_1 - 9.064736 \times 10^{-1}X_2 + 4.393847 \times 10^{-1}X_5$$

   Dari model diatas nilai R^2 yang diperoleh dari ketiga metode, yaitu sebesar 0.902 atau 90,2%, nilai $R^2_{Adjusted}$ sebesar 0.887 dan nilai MSE sebesar 0.113. Sehingga tidak terdapat perbedaan model regresi terbaik dengan metode *backward elimination, forward selection* dan *stepwise regression.*
3. Dari hasil model regresi linear dengan menggunakan metode *backward elimination, forward selection* dan *stepwise regression* diperoleh bahwa variabel bebas garis kemikinan, pendapatan perkapita dan jumlah penduduk miskin mempengaruhi persentase Penduduk Miskin di Provinsi Papua.
4. Persentase yang diperoleh dari metode *backward elimination, forward selection* dan *stepwise regression* adalah 90.2% dengan taraf signifikan yang dipilih 0.05 yang berarti bahwa variansi nilai Y dari persamaan regresi linear berganda sebesar 90.2%. Sisanya, sebesar 9.8% dijelaskan oleh faktor lain diluar variabel-variabel pada persamaan regresi diatas.

DAFTAR PUSTAKA

Chatterjee, S. & Hadi, A.S. 2006. *Regression Analysis by Example*. Ed. ke-4. Jhon Wiley & Sons Inc., New Jersey.

Drapper, N.R. and Smith, H. 1992. *Analisis Regresi Terapan.* Ed. ke-2. Diterjemahkan oleh Bambang Sumantri. Gramedia Pustaka Utama, Jakarta.

Freund, R. J., William, J., & Ping, S. 2006. *Regression Analysis: Statistical Modeling of a Response Variable*. Academic Press, Boston.

Gujarati, D. 1995. *Ekonometri Dasar.* Diterjemahkan oleh Sumarno Zain. Erlangga, Jakarta.

Kutner, M.H., et al. 2005. *Applied Linear Statistic Model.* ED. Ke-5. Mc-Graw-hill, New York.

Lindsey,C. dan Sheather, S. 2010. *Variable Selection in Linear Regression. The Stata.* **10**(2):650-669.

Montgomery, R.V., Peck, E.A & Vining, G. 2011. *Introduction to Linear Regression Analysis.* John Wiley & Sons, New York.

Myers, R.H. 1990. *Clasical and Modern Regression With Application.* PWSKENT publishing Company, Boston.

Myers, R.H. & Milton, J.S. 1991. *A First Course In The Theory Of Linier Statistical Models*. PWS-KENT Publishing Company, Boston.

Nirmalraj, J. dan Mallinga, P. 2011. *Airline Emission Forecast Through Empirical Analysis of Its Contributing Factors*. Internasional Journal of Environmental Science and Develpoment. **2**(1):ISSN 2010-0264.

Sembiring, R.K. 1995. *Analisis Regresi*. ITB, Bandung.

Sumodiningrat, G. 1998. *Ekonometrika Pengantar*. BPFE, Yogyakarta.

Usman, M. dan Warsono. 2000. *Teori Model Linear dan Aplikasinya*. C.V. Darmajaya, Bandar Lampung.