

# APLIKASI RANTAI MARKOV UNTUK MENENTUKAN PELUANG TRANSISI CURAH HUJAN

Iksan Kadafi, Sigit Nugroho, Pepi Novianti

Jurusan Matematika Fakultas MIPA Universitas Bengkulu

*Iksankadafi93@gmail.com*

## Abstrak

Rantai Markov merupakan salah satu metode yang digunakan untuk *forecasting* diberbagai bidang, seperti ekonomi, industri, dan iklim. Penelitian ini menggunakan data sekunder intensitas curah hujan harian dari BMKG Stasiun Geofisika Kabupaten Kepahiang. Penelitian ini bertujuan untuk menentukan peluang transisi (perpindahan) intensitas curah hujan harian, dimana ada empat *state* atau keadaan intensitas curah hujan yaitu tidak hujan, hujan ringan hujan sedang dan hujan lebat. Metode Rantai Markov yang digunakan adalah Persamaan Champman-Kolmogorov dan Persamaan *steady state*. Peluang tetap pada keadaan tidak hujan sebesar 58,90 % , peluang tetap pada keadaan hujan ringan sebesar 23,56 % , peluang tetap pada keadaan hujan sedang sebesar 13,15 % , dan peluang tetap pada keadaan hujan lebat sebesar 4,38 %.

Kata kunci: Rantai Markov, Curah Hujan

## PENDAHULUAN

Curah hujan merupakan salah satu unsur cuaca yang datanya diperoleh dengan cara menghitung jumlah air yang jatuh di permukaan tanah datar selama periode tertentu [1]. Curah hujan di Indonesia didominasi oleh adanya pengaruh beberapa fenomena, antara lain sistem Monsun Asia-Australia, Equatorial, El-Nino, sirkulasi Timur-Barat (*Walker Circulation*) dan Utara-Selatan (*Hadley Circulation*) serta beberapa sirkulasi karena pengaruh lokal [2].

Kondisi ketidak pastian curah hujan setiap harinya, karena mengalami perpindahan ke kondisi sama atau ke kondisi yang berbeda. Banyaknya kemungkinan perpindahan tersebut terjadi perlu diketahui. Oleh karena itu, perlu adanya teori untuk menentukan peluang curah hujan pada waktu akan datang. Salah satu teori yang sesuai dalam menentukan peluang curah hujan yang akan datang adalah Rantai Markov [3].

Rantai Markov merupakan sebuah proses stokastik, dimana kejadian pada

masa mendatang hanya tergantung pada kejadian hari ini dan tidak bergantung pada keadaan masa lampau. Rantai Markov terdefinisi oleh matriks peluang transisinya. Matriks peluang transisi adalah suatu matriks yang memuat informasi yang mengatur perpindahan sistem dari suatu keadaan (*state*) ke keadaan (*state*) lainnya langi [4].

Penelitian ini bertujuan untuk menerapkan metode Rantai Markov pada data curah hujan dan menentukan peluang untuk masing-masing *state* curah hujan.

## METODE PENELITIAN

### Data dan Variabel Penelitian

Data penelitian ini adalah data sekunder yang diperoleh dari Badan Meteorologi, Klimatologi, dan Geofiksa Stasiun Geofisika Kabupaten Kepahiang, Provinsi Bengkulu. Data tersebut adalah data curah hujan harian dari tanggal 1 Januari 2016 sampai 31 Desember 2016.

Variabel Penelitian yaitu

$$X_t = \begin{cases} \text{tidak hujan} & \text{untuk curah hujan } < 5 \text{ mm} \\ \text{hujan ringan} & \text{untuk curah hujan } 5\text{-}20 \text{ mm} \\ \text{hujan sedang} & \text{untuk curah hujan } 20\text{-}50 \text{ mm} \\ \text{hujan lebat} & \text{untuk curah hujan } 50\text{-}100 \text{ mm} \end{cases}$$

$$X_0 = i, X_1 = i, X_2 = i, \dots, X_{366} = i$$

### Tahapan Penelitian

Membuat tabel perpindahan (transisi) curah hujan dari hari pertama sampai ke hari berikutnya sampai pada ke tiga puluh  $X_0, X_1, X_2, \dots, X_{366}$ , sesuai dengan

ketentuan sifat Rantai Markov yaitu hari ke tiga hanya tergantung pada hari ke dua, tidak tergantung dengan hari pertama  $\{X_{t+1} = j | X_t = i\}$  untuk  $t = 0, 1, 2, \dots, 366$

dan  $i, j = 0, 1, 2, 3$ . Menentukan peluang transisi  $P_{ij}$  dari tabel perpindahan curah

hujan setiap cell diganti dengan nilai peluang, dengan cara jumlah perpindahan cell yang ingin diganti dibagi dengan total jumlah perpindahan cell pada baris yang ingin diganti. Hal tersebut dilakukan untuk memenuhi sifat dari peluang transisi

$\sum_{j=0}^3 P_{ij} = 1, i = 0, 1, 2, 3$ . Membuat matriks

peluang transisi satu langkah  $\mathbf{P} = [P_{ij}]$

Menggambarkan diagram transisi matriks peluang transisi satu langkah. Menentukan matriks peluang transisi  $n$  langkah  $\mathbf{P}^{(n)}$  dari matriks peluang transisi satu langkah dengan Persamaan Chapman-Kolmogorov. Menentukan peluang steady-state dengan menggunakan Persamaan steady-state, selanjutnya menyimpulkan peluang transisi curah hujan pada kabupaten Kepahiang yang akan datang

### HASIL DAN PEMBAHASAN

Pada penelitian ini data yang digunakan untuk dianalisis adalah data curah hujan harian stasiun Geofisika Kabupaten Kepahiang selama setahun mulai dari 1 Januari sampai 31 Desember 2016 sebanyak 366 hari. Data curah hujan yang diperoleh memiliki curah hujan

terendah 0 mm dan 90,2 mm adalah curah hujan tertinggi dengan rata-rata perhari 10,0637 mm. Informasi yang didapatkan dari angka curah hujan setiap harinya diubah menjadi keterangan atau dinamakan intensitas curah hujan. Adapun pengelompokkan untuk intensitas telah ditetapkan oleh BMKG, dimana *state* 0 untuk tidak hujan yang memiliki curah hujan kurang dari 5 mm per hari, *state* 1 untuk hujan ringan yang memiliki curah hujan 5 sampai 20 mm per hari, *state* 2 untuk hujan sedang yang memiliki curah hujan 20 sampai 50 mm per hari, dan *state* 3 untuk hujan lebat yang memiliki curah hujan 50 sampai 100 mm per hari. Pada tahun 2016 intensitas curah hujan terjadi adalah

**Tabel 4.1 Intensitas Curah Hujan**

Intensitas Curah Hujan	Banyak (hari)
Tidak Hujan	216
Hujan Ringan	86
Hujan Sedang	48
Hujan Lebat	16
Jumlah	366

Berdasarkan tabel di atas terjadi selama setahun keadaan tidak hujan sebanyak 216 hari, keadaan hujan ringan sebanyak 86 hari, keadaan hujan sedang sebanyak 48 hari, dan keadaan hujan lebat sebanyak 16 hari.

Intensitas hujan yang terjadi dari hari ke hari selama tahun 2016 mengalami perubahan yang tidak menentu, ada yang tetap pada *state* semula dan ada juga yang berubah ke *state* yang lain. Pada tahun 2016 perpindahan yang terjadi dapat dilihat pada Tabel 4.2

**Tabel 4.2 Jumlah transisi state**

State	0	1	2	3	Total
0	131	46	27	11	215
1	56	17	12	1	86
2	20	16	8	4	48
3	8	7	1	0	16
Jumlah					365

Pada *state* hari ini tidak hujan berpindah ke *state* besok yang sama yaitu tidak hujan terjadi sebanyak 131 hari, perpindahan yang sama atau tetap pada *state* hujan ringan ke hujan ringan sebanyak 17 hari, begitupula terjadi pada *state* hujan sedang yaitu sebanyak 8 hari, dan tidak terjadi *state* hujan lebat ke hujan lebat. Adapun *state* hari ini tidak hujan berpindah ke *state* besok hujan ringan, hujan sedang, dan hujan lebat masing-masing adalah 46 hari, 27 hari, dan 11 hari, sedangkan *state* hari ini hujan ringan ke *state* besok tidak hujan, hujan sedang, dan hujan lebat yaitu sebanyak 56 hari, 12 hari, dan 1 hari. *State* hari ini hujan sedang berpindah ke *state* besok tidak hujan, hujan ringan, dan hujan lebat masing-masing sebanyak 20 hari, 16 hari, dan 4 hari, kemudian untuk *state* hari ini hujan lebat ke *state* besok tidak hujan, hujan ringan, dan hujan sedang yaitu sebanyak 8 hari, sebanyak 7 hari, dan sebanyak 1 hari.

Proses dalam mengetahui jumlah perpindahan *state* atau dengan kata lain transisi berdasarkan sifat Rantai Markov pada Persamaan 2.1 dan perhitungannya dibantu oleh Microsoft Excel. Hal tersebut dapat dilihat pada Lampiran 1. bagian transisi. Tabel 4.2 diubah ke dalam bentuk peluang dengan cara sebagai berikut

**Tabel. 4.3 Peluang transisi *state***

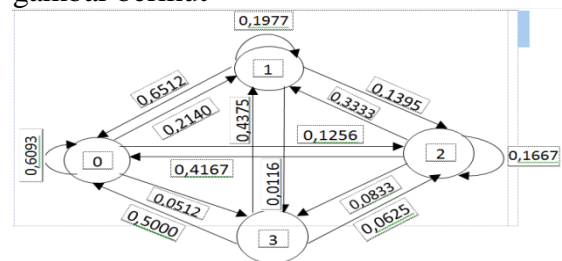
<i>State</i>	0	1	2	3	Total Peluang
0	$\frac{131}{215}$	$\frac{46}{215}$	$\frac{27}{215}$	$\frac{11}{215}$	$\frac{215}{215} = 1$
1	$\frac{56}{86}$	$\frac{17}{86}$	$\frac{12}{86}$	$\frac{1}{86}$	$\frac{86}{86} = 1$
2	$\frac{20}{48}$	$\frac{16}{48}$	$\frac{8}{48}$	$\frac{4}{48}$	$\frac{48}{48} = 1$
3	$\frac{8}{16}$	$\frac{7}{16}$	$\frac{1}{16}$	$\frac{0}{16}$	$\frac{16}{16} = 1$

Nilai peluang dari Tabel 4.3 dapat disajikan/ dimasukkan ke dalam sebuah matriks berukuran 4X4 disebut sebagai

matriks peluang transisi satu langkah dan sebagai berikut

$$P^{(1)} = \begin{bmatrix} 0,6093 & 0,2140 & 0,1256 & 0,0512 \\ 0,6512 & 0,1977 & 0,1395 & 0,0116 \\ 0,4167 & 0,3333 & 0,1667 & 0,0833 \\ 0,5000 & 0,4375 & 0,0625 & 0 \end{bmatrix}$$

Matriks Peluang transisi satu langkah bisa disajikan juga dalam diagram transisi, dimana tanda panah menunjukkan transisi dan lingkaran adalah *state* pada gambar berikut



Persamaan Chapman-Kolmogorov digunakan untuk mencari nilai matriks peluang transisi beberapa langkah, sebagai berikut  
 Persamaan Chapman-Kolmogorov digunakan untuk mencari nilai matriks peluang transisi beberapa langkah, sebagai berikut

$$P^{(1)} = \begin{bmatrix} 0,6093 & 0,2140 & 0,1256 & 0,0512 \\ 0,6512 & 0,1977 & 0,1395 & 0,0116 \\ 0,4167 & 0,3333 & 0,1667 & 0,0833 \\ 0,5000 & 0,4375 & 0,0625 & 0 \end{bmatrix}$$

$$P^{(2)} = P^{(1)} \cdot P^{(1)} = \begin{bmatrix} 0,5885 & 0,2369 & 0,1305 & 0,0441 \\ 0,5894 & 0,2300 & 0,1333 & 0,0472 \\ 0,5820 & 0,2471 & 0,1318 & 0,0391 \\ 0,6156 & 0,2143 & 0,1343 & 0,0359 \end{bmatrix}$$

⋮

$$P^{(6)} = \begin{bmatrix} 0,5890 & 0,2356 & 0,1315 & 0,0438 \\ 0,5890 & 0,2356 & 0,1315 & 0,0438 \\ 0,5890 & 0,2356 & 0,1315 & 0,0438 \\ 0,5890 & 0,2356 & 0,1315 & 0,0438 \end{bmatrix}$$

Berdasarkan perkalian matriks tersebut menunjukkan bahwa matriks peluang transisi pada langkah 1 (pada tahun 2016) menuju langkah 6 (pada tahun 2021) atau  $P^{(6)}$  nilai peluangnya mengalami perubahan atau berbeda, namun pada langkah ke 7 dan seterusnya nilai matriks peluang transisi sama dengan nilai  $P^{(6)}$ . Oleh karena itu, pada tahun 2021 hingga tahun berikutnya peluang transisinya sama atau tetap.

Metode *Steady state* digunakan untuk mencari nilai peluang tetap untuk setiap *state*. Penyelesaian secara dengan menggunakan metode eliminasi dan substitusi dan dengan bantuan *Microsoft Excell.* metode eliminasi dan substitusi untuk mencari nilai  $\pi_0, \pi_1, \pi_2$ , dan  $\pi_3$  dan peluang awal. Tulis

$$[\pi_0, \pi_1, \pi_2, \pi_3] = [\pi_0 \quad \pi_1 \quad \pi_2 \quad \pi_3] \begin{bmatrix} P_{00} & P_{01} & P_{02} & P_{03} \\ P_{10} & P_{11} & P_{12} & P_{13} \\ P_{20} & P_{21} & P_{22} & P_{23} \\ P_{30} & P_{31} & P_{32} & P_{33} \end{bmatrix}$$

dan  $\pi_0 + \pi_1 + \pi_2 + \pi_3 = 1$

$$[\pi_0, \pi_1, \pi_2, \pi_3] = [\pi_0 \quad \pi_1 \quad \pi_2 \quad \pi_3] \begin{bmatrix} 0,6093 & 0,2140 & 0,1256 & 0,0512 \\ 0,6512 & 0,1977 & 0,1395 & 0,0116 \\ 0,4167 & 0,3333 & 0,1667 & 0,0833 \\ 0,5000 & 0,4375 & 0,0625 & 0 \end{bmatrix}$$

dari perkalian matriks di atas diperoleh

$$\begin{aligned} (\pi_0 - 0,6093\pi_0) - 0,6512\pi_1 - 0,4167\pi_2 - 0,500\pi_3 &= 0 \\ 0,2140\pi_0 + (-\pi_1 + 0,1977\pi_1) + 0,3333\pi_2 + 0,4375\pi_3 &= 0 \\ 0,1256\pi_0 + 0,1395\pi_1 + (-\pi_2 + 0,1667\pi_2) + 0,0625\pi_3 &= 0 \\ 0,0512\pi_0 + 0,0116\pi_1 + 0,0833\pi_2 - \pi_3 &= 0 \\ \pi_0 + \pi_1 + \pi_2 + \pi_3 &= 1 \end{aligned}$$

dapat disederhanakan menjadi

$$\begin{aligned} 0,3907\pi_0 - 0,6512\pi_1 - 0,4167\pi_2 - 0,500\pi_3 &= 0 \\ 0,2140\pi_0 - 0,8023\pi_1 + 0,3333\pi_2 + 0,4375\pi_3 &= 0 \\ 0,1256\pi_0 + 0,1395\pi_1 - 0,8333\pi_2 + 0,0625\pi_3 &= 0 \\ 0,0512\pi_0 + 0,0116\pi_1 + 0,0833\pi_2 - 1\pi_3 &= 0 \\ \pi_0 + \pi_1 + \pi_2 + \pi_3 &= 1 \end{aligned}$$

Hasil yang diperoleh dari metode *steady state* di atas nilai peluang transisi tetap masing-masing *state* adalah

$$[\pi_0, \pi_1, \pi_2, \pi_3] = [0,5890, 0,2356, 0,1315, 0,0348]$$

## KESIMPULAN

Berdasarkan dari hasil dan pembahasan penelitian dapat disimpulkan:

Metode Rantai Markov dapat diterapkan untuk menentukan peluang perpindahan intensitas curah hujan harian. Peluang masing-masing *state* untuk jangka waktu kedepan yang panjang adalah 58,90% untuk peluang tetap pada keadaan tidak hujan, 23,56% untuk peluang tetap pada keadaan hujan ringan, 13,15% untuk peluang tetap pada keadaan hujan sedang, dan 4,38% untuk peluang tetap pada keadaan hujan lebat.

## DAFTAR PUSTAKA

- [1] Prawaka, F., Z.A., and S., Tugiono, 2016, Analisis Data Curah Hujan yang Hilang dengan menggunakan Metode Normal *Ratio*, *Inversed Square Distance*, dan Rata-Rata Aljabar (Studi kasus Curah Hujan Beberapa Stasiun Hujan Daerah Bandar Lampung), *Jurnal RSDD*, No.3, Vol. 4, Hal.397-406.
- [2] Affandi, R., A., Lubis dan D., Septiadi, 2012, Karakteristik Pola Curah Hujan di Wilayah Sekitar Teluk (Studi Daerah Nabire Provinsi Papua dan Fakfak Papua Barat), *Jurnal Matematika dan Sains*, No.2, Vol.17, Hal.47-54.
- [3] Pratiwi, R. Y., 2012, Pemodelan Curah Hujan dengan Campuran Rantai Markov dan Model Deret Waktu, *Jurnal Matematika FMIPA Universitas Brawijaya*.
- [4] Langi, Y. A. R., 2011, Penentuan Klasifikasi *State* pada Rantai Markov dengan menggunakan Nilai *Eigen* dari Matriks Peluang Transisi, *Jurnal ilmiah Sains*, No.1, Vol.11, Hal.124.

# METODE *LATENT ROOT REGRESSION* (LRR) DALAM MENGATASI MASALAH MULTIKOLINIERITAS

Neni Agustini<sup>1</sup>, Sigit Nugroho<sup>2</sup> dan Dian Agustina<sup>3</sup>

<sup>1</sup>Alumni Jurusan Matematika, FMIPA, Universitas Bengkulu

<sup>2,3</sup>Staf Pengajar Jurusan Matematika, FMIPA, Universitas Bengkulu

e-mail: [1neni.agustini@gmail.com](mailto:1neni.agustini@gmail.com), [2sigit.nugroho.1960@gmail.com](mailto:2sigit.nugroho.1960@gmail.com) dan [3dianagustina117@yahoo.com](mailto:3dianagustina117@yahoo.com)

## ABSTRAK

Penelitian ini bertujuan untuk mengetahui dan memahami cara mendeteksi adanya multikolinieritas dan prosedur metode *Latent Root Regression* (LRR) dalam mengatasi masalah multikolinieritas. Agar lebih mudah memahami metode LRR, maka perlu teladan penerapan. Teladan penerapan diambil dari buku “*Applied Linear Statistical Models Fifth Edition*” oleh Kutner, *et al.* (2005: 256-257) tentang persentase lemak tubuh wanita berusia 25-34 tahun dengan variabel yang mempengaruhinya. Hasil analisis dalam penelitian ini diperoleh bahwa metode LRR menghasilkan nilai *VIF* (*Variance Inflation Factor*) yang tidak lagi melebihi 10 dan tidak ada korelasi antar variabel bebas. Selain itu nilai koefisien determinasi yang dihasilkan sebesar 100%. Sehingga dapat dikatakan bahwa metode LRR menunjukkan masalah multikolinieritas dapat diatasi secara tuntas.

**Kata Kunci:** Multikolinieritas, Korelasi, *Variance Inflation Factor* (VIF), *Latent Root Regression* (LRR)

## 1. PENDAHULUAN

Analisis regresi adalah salah satu metode atau teknik statistika yang dapat digunakan untuk menyelidiki hubungan atau pengaruh antara suatu variabel dengan variabel lainnya. Variabel-variabel regresi yang berhubungan secara linier disebut regresi linier. Regresi linier yang menghubungkan satu variabel respon dengan satu variabel bebas disebut dengan regresi linier sederhana, sedangkan regresi linier yang menghubungkan satu variabel respon dengan dua atau lebih variabel bebas disebut regresi linier berganda (Kutner, *et al.*, 2004).

Salah satu asumsi yang harus dipenuhi dalam model regresi linier berganda adalah tidak terdapat multikolinieritas di antara variabel bebas yang termasuk dalam model. Multikolinieritas merupakan suatu kondisi dimana terjadi korelasi yang tinggi antar variabel-variabel bebas yang mengakibatkan determinan dari matriks  $X'X$  akan mendekati 0 sehingga akan menyebabkan matriks tersebut hampir *singular* yang mengakibatkan nilai dari penduga parameternya tidak dapat ditemukan (Hoerl dan Kennard, 1970).

Untuk mengetahui adanya multikolinieritas yaitu dengan melihat nilai  $VIF > 10$ . Gejala multikolinieritas menimbulkan masalah dalam model regresi. Korelasi antar

<sup>1</sup>Alumni Jurusan Matematika FMIPA Universitas Bengkulu

<sup>2,3</sup>Staf Pengajar Jurusan Matematika FMIPA Universitas Bengkulu

variabel bebas yang sangat tinggi menghasilkan penduga model regresi yang berbias, tidak stabil, dan mungkin jauh dari nilai prediksinya (Bilfarsah, 2005).

Beberapa penelitian mengenai metode yang dapat digunakan untuk mengatasi masalah multikolinieritas telah banyak dilakukan, seperti metode regresi ridge, *partial least square*, analisis faktor, regresi komponen utama dan *latent root regression* (LRR). Dalam hal ini, metode *Latent Root Regression* (LRR) belum banyak yang menggunakannya dalam sebuah penelitian. Oleh karena itu, perlu adanya pembahasan lebih lanjut mengenai *Latent Root Regression* (LRR).

Metode *Latent Root Regression* (LRR) merupakan perluasan dari metode regresi komponen utama. Perluasan regresi komponen utama untuk pemeriksaan persamaan peramalan alternatif dan untuk pembuangan peubah peramal yang telah diajukan oleh J. T. Webster, R. F. Gunst, dan R. L. Mason, dalam "*Latent root regression analysis*", *Technometrics*, 16, 1974. Webster dan rekannya menggabungkan matriks data yang berasal dari variabel peramal dengan variabel respon yang telah dibakukan. Perluasan ini dinamakan *Latent Root Regression* (Draper dan Smith, 1992).

Konsep regresi komponen utama sendiri yaitu dengan tahap analisis dengan menggunakan analisis komponen utama. Melalui penggunaan analisis ini akan dihasilkan variabel-variabel baru (komponen utama) yang merupakan kombinasi linier dari variabel-variabel bebas asal dan antar variabel bebas baru ini bersifat saling bebas. Selanjutnya, diregresikan dengan variabel respon (Marcus, Wattimanela dan Lesnussa, 2012). Sedangkan komponen utama dalam regresi akar laten diperoleh dari matriks korelasi gandingan dengan variabel bebas dan variabel respon yang telah dibakukan.

Berdasarkan penelitian dan uraian di atas, maka penulis tertarik untuk membahas masalah multikolinieritas dan metode yang dapat mengatasinya, yaitu menggunakan metode *Latent Root Regression* (LRR) dalam mengatasi masalah multikolinieritas.

## 2. LANDASAN TEORI

### 2.1 Matriks

Secara umum, sebuah matriks  $A$  yang berukuran  $m$  baris dan  $n$  kolom dengan elemen  $a_{ij}$  dapat dituliskan sebagai berikut:

$$A_{m \times n} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \quad (2.1)$$

dengan  $i = 1, 2, \dots, m$  dan  $j = 1, 2, \dots, n$ .

### 2.2 Determinan matriks

**Definisi 2.1** (Santosa, 2009:25)

Bila matriks  $A$  adalah matriks berukuran  $n \times n$ , maka minor elemen  $a_{ij}$  (disimbolkan dengan  $M_{ij}$ ) didefinisikan sebagai determinan dari submatriks yang ada setelah baris ke- $i$  dan kolom ke- $j$  dicoret dari  $A$ . Nilai  $(-1)^{i+j}$  dituliskan sebagai  $C_{ij}$  dan dinamakan **kofaktor** elemen  $a_{ij}$ . Jadi,  $C_{ij} = (-1)^{i+j}M_{ij}$ .

### 2.3 Nilai eigen dan vektor eigen

**Definisi 2.2** (Santosa, 2009:159)

Jika  $A$  adalah sebuah matriks berukuran  $n \times n$ , maka sebuah vektor tak nol  $\gamma$  di  $R^n$  dinamakan vektor eigen dari  $A$  jika  $A\gamma$  adalah kelipatan skalar dari  $\gamma$ , yaitu:

$$A\gamma = \lambda\gamma \quad (2.2)$$

<sup>1</sup>Alumni Jurusan Matematika FMIPA Universitas Bengkulu

<sup>2,3</sup>Staf Pengajar Jurusan Matematika FMIPA Universitas Bengkulu

Skalar  $\lambda$  ini dinamakan **nilai eigen** dari  $A$ , sedangkan  $\gamma$  dinamakan **vektor eigen** yang bersesuaian dengan  $\lambda$ .

#### 2.4 Analisis regresi linier berganda

Model pada analisis regresi linier berganda adalah:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} + \varepsilon_i \quad (2.3)$$

Keterangan:

- $Y_i$  : variabel respon pengamatan ke- $i$
- $X_{ij}$  : variabel bebas ke- $j$  pada pengamatan ke- $i$
- $\beta_0, \beta_1, \dots, \beta_k$  : parameter regresi
- $\varepsilon_i$  : galat (*error*) pengamatan ke- $i$  dan  $\varepsilon_i \sim N(0, \sigma^2)$
- $i = 1, 2, \dots, n$  dan  $j = 1, 2, \dots, k$

##### 2.4.1 Asumsi regresi linier berganda

Menurut Gujarati (2003) asumsi-asumsi pada model regresi linier berganda adalah sebagai berikut:

1. *Error* berdistribusi normal
2. Variansi dari *error* adalah konstan (homoskedastisitas)
3. Tidak terjadi autokorelasi pada *error* (untuk data time series)
4. Tidak terjadi multikolinieritas pada variabel bebas

##### 2.4.2 Estimasi parameter

Menurut Kutner, *et al* (2005) salah satu metode yang digunakan dalam mengestimasi parameter adalah metode kuadrat terkecil atau sering juga disebut dengan metode *Ordinary Least Square* (OLS). Metode OLS ini bertujuan untuk mengestimasi parameter dengan cara meminimumkan jumlah kuadrat galat:

$$Q = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_{i1} - \beta_2 X_{i2} - \dots - \beta_k X_{ik})^2 \quad (2.4)$$

Sehingga diperoleh penduga kuadrat terkecil dari  $\beta$  sebagai berikut:

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \quad (2.5)$$

##### 2.4.3 Pengujian parameter

1. Pengujian parameter secara simultan

Hipotesis:

$H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$  Vs  $H_1$  : paling sedikit ada satu  $\beta_j \neq 0$  dengan  $j = 1, 2, \dots, k$

Statistik uji:

$$F_{hitung} = \frac{KTR}{KTG} = \frac{JKR / (k - 1)}{JKG / (n - k)} = \frac{JKR(n - k)}{JKG(k - 1)} \quad (2.6)$$

Titik kritis:

$$F_{tabel} = F_{(1-\alpha; k, (n-k-1))}$$

Daerah penolakan:

$H_0$  ditolak pada taraf  $\alpha$  jika  $F_{hitung} > F_{tabel}$ , sebaliknya jika  $F_{hitung} \leq F_{tabel}$  maka  $H_0$  diterima pada taraf  $\alpha$ .

<sup>1</sup>Alumni Jurusan Matematika FMIPA Universitas Bengkulu

<sup>2,3</sup>Staf Pengajar Jurusan Matematika FMIPA Universitas Bengkulu

## 2. Pengujian parameter secara parsial

Hipotesis:

$$H_0 : \beta_j = 0 \text{ Vs } H_1 : \beta_j \neq 0 \text{ dengan } j = 1, 2, \dots, k$$

Statistik uji:

$$t = \frac{b_j}{s\{b_j\}} \quad (2.7)$$

Titik Kritis:

$$t_{tabel} = t_{(1-\alpha/2; (n-k-1))}$$

Daerah penolakan:

Jika  $t > t_{tabel}$  untuk derajat bebas  $n - k - 1$  maka hipotesis  $H_0$  ditolak pada taraf  $\alpha$ , sebaliknya jika  $t \leq t_{tabel}$  maka  $H_0$  diterima pada taraf  $\alpha$ .

## 2.5 Korelasi

Koefisien korelasi adalah angka yang menunjukkan tingginya derajat hubungan antara dua variabel acak yang diteliti dan merupakan nilai baku dari kovarian kedua variabel tersebut. Nilai korelasi terletak di antara -1 dan 1 ( $-1 \leq r \leq 1$ ) (Soelistyo, 2000).

## 2.6 Koefisien determinasi

Proporsi keragaman data yang dapat dijelaskan dalam model regresi dilihat dari koefisien determinasi yang dilambangkan dengan  $R^2$ . Koefisien determinasi didefinisikan sebagai berikut:

$$R^2 = \frac{b'X'Y - \left(\frac{1}{n}\right)Y'JY}{Y'Y - \left(\frac{1}{n}\right)Y'JY} = \frac{JKR}{JKT} \quad (2.8)$$

Nilai  $R^2$  terletak pada interval  $0 \leq R^2 \leq 1$ .

## 2.7 Multikolinieritas

Istilah multikolinieritas pertama kali diperkenalkan oleh Ragnar Frisch pada tahun 1934, yang berarti adanya korelasi di antara variabel-variabel bebas dari model regresi. Multikolinieritas adalah terjadinya hubungan linier antara variabel bebas dalam suatu model regresi linier berganda (Gujarati, 2003).

Untuk mendeteksi adanya multikolinieritas dapat dilihat nilai *Variance Inflation Factor* (VIF) dan *Tolerance* (TOL) dengan ketentuan jika nilai  $VIF > 10$ , maka terjadi multikolinieritas dalam model regresi. VIF merupakan elemen diagonal utama dari matriks korelasi  $((X'X)^{-1})$ .

## 2.8 Regresi komponen utama

Regresi komponen utama merupakan teknik analisis regresi yang dikombinasikan dengan analisis komponen utama yang dijadikan sebagai tahap analisis (Marcus, Wattimanela dan Lenussa, 2012).

Tahap pertama pada regresi komponen utama adalah menghitung komponen utama yang merupakan kombinasi linier dari variabel bebas. Langkah selanjutnya, beberapa komponen utama yang terbentuk diregresikan dengan variabel respon melalui analisis regresi (Myers dan Milton, 1991). Kriteria pemilihan komponen utama yang akan digunakan yaitu dengan memilih komponen utama yang bersesuaian dengan akar ciri lebih besar dari 1 (Draper dan Smith, 1992).

<sup>1</sup>Alumni Jurusan Matematika FMIPA Universitas Bengkulu

<sup>2,3</sup>Staf Pengajar Jurusan Matematika FMIPA Universitas Bengkulu

### 3. METODE LATENT ROOT REGRESSION (LRR)

#### 3.1 Pemusatan dan penskalaan

Pemusatan menjelaskan perbedaan antara masing-masing pengamatan dan rata-rata dari semua pengamatan untuk variabel. Sedangkan penskalaan menjelaskan gambaran pemusatan pengamatan pada kesatuan (unit) standar deviasi dari pengamatan untuk variabel (Kutner, *et al.*, 2005).

Pembakuan variabel respon  $Y$  dan variabel bebas  $X_1, X_2, \dots, X_k$  dalam hal ini dapat dilakukan sebagai berikut:

$$\frac{Y_i - \bar{Y}}{S_Y}, \text{ di mana } S_Y = \sqrt{\frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n-1}} \quad (3.1)$$

$$\frac{X_{ij} - \bar{X}_j}{S_{X_j}}, \text{ di mana } S_{X_j} = \sqrt{\frac{\sum_{i=1}^n (X_{ij} - \bar{X}_j)^2}{n-1}} \quad (3.2)$$

Keterangan:

$\bar{Y}$  : rata-rata dari  $Y$

$\bar{X}_j$  : rata-rata dari pengamatan  $X_j$

$S_Y$  : standar deviasi dari  $Y$

$S_{X_j}$  : standar deviasi dari  $X_j$

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

Sebelumnya telah dijelaskan bahwa transformasi korelasi merupakan fungsi sederhana dari pembakuan variabel pada persamaan (3.1) dan (3.2). Sehingga dengan cara transformasi korelasi maka diperoleh:

$$\mathbf{Z}_y = [Z_{iy}] = \left[ \frac{1}{\sqrt{n-1}} \left( \frac{Y_i - \bar{Y}}{S_Y} \right) \right]; \quad i = 1, 2, \dots, n \quad (3.3)$$

$$\mathbf{Z} = [Z_{ij}] = \left[ \frac{1}{\sqrt{n-1}} \left( \frac{X_{ij} - \bar{X}_j}{S_{X_j}} \right) \right]; \quad i = 1, 2, \dots, n \text{ dan } j = 1, 2, \dots, k \quad (3.4)$$

Berdasarkan variabel yang ditransformasi  $Z_{iy}$  dan  $Z_{ij}$  yang didefinisikan dengan transformasi korelasi pada persamaan (3.3) dan (3.4) di atas, maka diperoleh model regresi sebagai berikut:

$$Z_{iy} = \beta_1^* Z_{i1} + \beta_2^* Z_{i2} + \dots + \beta_k^* Z_{ik} + \varepsilon_i^* \quad (3.5)$$

Persamaan (3.5) di atas disebut sebagai model regresi yang baku (*standardized regression model*).

#### 3.2 Matriks korelasi gabungan

Menurut Draper dan Smith (1992) pada metode LRR, matriks korelasi yang digunakan merupakan matriks yang diperoleh dari penggabungan antara variabel respon dan variabel bebas yang telah dibakukan dengan  $\mathbf{Z}_y$  dan  $\mathbf{Z}$  dapat ditulis dalam bentuk matriks, yaitu:

$$\mathbf{Z}^* = \begin{bmatrix} Z_{1y} & Z_{11} & \dots & Z_{1k} \\ Z_{2y} & Z_{21} & \dots & Z_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ Z_{ny} & Z_{n1} & \dots & Z_{nk} \end{bmatrix} \quad (3.6)$$

Berdasarkan persamaan (3.6) maka matriks korelasi gabungannya dapat dituliskan sebagai berikut:

<sup>1</sup>Alumni Jurusan Matematika FMIPA Universitas Bengkulu

<sup>2,3</sup>Staf Pengajar Jurusan Matematika FMIPA Universitas Bengkulu

$$\mathbf{Z}^* \mathbf{Z}' = \begin{bmatrix} 1 & r_{2Y} & \cdots & r_{kY} \\ r_{2Y} & 1 & \cdots & r_{k1} \\ \vdots & \vdots & \ddots & \vdots \\ r_{kY} & r_{k1} & \cdots & 1 \end{bmatrix} \quad (3.7)$$

### 3.3 Analisis komponen utama pada metode LRR

Seperti halnya dalam analisis komponen utama, akar laten dan vektor laten dapat dihitung dari matriks korelasi gabungan  $\mathbf{Z}^* \mathbf{Z}'$ . Menurut Webster, Gunst dan Mason (1974) dengan mengikuti definisi 2.10 serta persamaan (2.3) dan (2.4) akar laten dan vektor laten dapat didefinisikan sebagai berikut:

$$|\mathbf{Z}^* \mathbf{Z}' - \lambda \mathbf{I}| = 0 \quad (3.8)$$

dan

$$(\mathbf{Z}^* \mathbf{Z}' - \lambda_j \mathbf{I}) \gamma_j = \mathbf{0} \quad ; j = 0, 1, 2, \dots, k \quad (3.9)$$

Misalkan  $\gamma_j' = (\gamma_{0j}, \gamma_{1j}, \gamma_{2j}, \dots, \gamma_{kj})$  merupakan vektor laten ke- $j$  dari matriks  $\mathbf{Z}^* \mathbf{Z}'$  dan  $\gamma_j^0 = (\gamma_{1j}, \gamma_{2j}, \dots, \gamma_{kj})$  merupakan vektor laten yang terbentuk dari elemen yang sama dengan  $\gamma_j'$  kecuali elemen pertama yang telah dibuang. Akhirnya, misalkan  $\Gamma = (\gamma_0, \gamma_1, \dots, \gamma_k)$  dan  $\Lambda = \text{diag}(\lambda_0, \lambda_1, \dots, \lambda_k)$ , maka  $\Gamma' (\mathbf{Z}^* \mathbf{Z}') \Gamma = \Lambda$  dan  $\mathbf{Z}^* \mathbf{Z}' = \Gamma \Lambda \Gamma'$ .

Selanjutnya, komponen utama (*principal component*) dari matriks  $\mathbf{Z}^*$  adalah:

$$PC_j = \mathbf{Z}_y \gamma_{0j} + \mathbf{Z} \gamma_j^0 \quad (3.10)$$

dan

$$\lambda_j = \sum_{i=1}^n \left( Z_{iy} * \gamma_{0j} + \sum_{r=1}^k Z_{ir} * \gamma_{rj} \right)^2 \quad (3.11)$$

Selanjutnya, komponen utama yang akan digunakan pada tahap analisis diperoleh dengan membuang komponen utama yang bersesuaian dengan nilai akar laten  $\lambda_j \leq 0,05$  dan elemen pertama vektor laten  $|\gamma_{0j}| < 0,10$  (Webster, Gunst dan Mason, 1974).

### 3.4 Koefisien regresi termodifikasi

Sehingga diperoleh koefisien kuadrat terkecil termodifikasi yaitu

$$\mathbf{b}^* = \begin{bmatrix} b_1^* \\ b_2^* \\ \vdots \\ b_k^* \end{bmatrix} = c \sum_{j=p}^k \gamma_{0j} \lambda_j^{-1} \begin{bmatrix} \gamma_{1j} \\ \gamma_{2j} \\ \vdots \\ \gamma_{rj} \end{bmatrix} \quad (3.12)$$

dengan  $c$  adalah konstanta yaitu:

$$c = - \left\{ \sum_{j=p}^k \gamma_{0j}^2 \lambda_j^{-1} \right\}^{-1} \left\{ \sum_{i=1}^n (Y_i - \bar{Y})^2 \right\}^{1/2} \quad (3.13)$$

Keterangan:

$\lambda_j$  : akar laten ke- $j$  dari matriks  $\mathbf{Z}^* \mathbf{Z}'$

$\gamma_{tj}$  : elemen vektor laten ke- $j$

$\gamma_{0j}$  : elemen pertama dari vektor laten ke- $j$

$j = 1, 2, \dots, k$  dan  $t = 1, 2, \dots, r$

<sup>1</sup>Alumni Jurusan Matematika FMIPA Universitas Bengkulu

<sup>2,3</sup>Staf Pengajar Jurusan Matematika FMIPA Universitas Bengkulu

Selanjutnya, menurut Draper dan Smith (1992) penduga koefisien regresi pada variabel asal diperoleh dengan membagi penduga koefisien regresi pada variabel yang telah dibakukan dengan  $S_j$  sehingga diperoleh:

$$b_j = \frac{b_j^*}{S_j} \quad (3.14)$$

dengan

$$S_j = \sqrt{\sum_{j=1}^k (X_j - \bar{X}_j)^2}$$

Sedangkan untuk perhitungan koefisien regresi  $b_0$  diperoleh berdasarkan rumus:

$$b_0 = \bar{Y} - b_1\bar{X}_1 - b_2\bar{X}_2 - \dots - b_k\bar{X}_k \quad (3.15)$$

### 3.5 Metode eliminasi *backward*

Menurut Webster, Gunst dan Mason (1974) kemudian untuk menentukan variabel bebas yang akan dihapus dari model maka dapat digunakan statistik uji- $F$  yaitu

$$F = \left[ \frac{SSE(\text{semua } X \text{ kecuali } X_r)}{\eta^2 \rho_{00}^{-1}} \right] (n - k - 1) \quad (3.16)$$

sehingga menjadi

$$F = \left[ \frac{\eta^2 \left( \rho_{00} - \frac{\rho_{0r}^2}{\rho_{rr}} \right)^{-1}}{\eta^2 \rho_{00}^{-1}} \right] (n - k - 1) \quad (3.17)$$

dengan pembandingan  $F_{tabel} = F_{(1-\alpha; 1, (n-k-1))}$ . Jika  $F_{hitung} > F_{tabel}$  untuk derajat bebas 1 dan  $n - k - 1$  maka variabel bebas  $X_r$  tidak dihapus dari model. Sebaliknya, jika  $F_{hitung} \leq F_{tabel}$  maka variabel bebas  $X_r$  dihapus dari model.

## 4. TELADAN PENERAPAN

Teladan yang digunakan pada metode *Latent Root Regression* (LRR) dalam mengatasi masalah multikolinieritas diambil dari teladan pada buku "*Applied Linear Statistical Models Fifth Edition*" oleh Kutner, *et al.* (2005: 256-257). Data disajikan dalam Tabel 1 mengenai data persentase lemak tubuh 20 wanita dengan variabel yang mempengaruhinya yaitu *triceps skinfold thickness* (mm) ( $X_1$ ), *thigh circumference* (mm) ( $X_2$ ) dan *mid arm circumference* (mm) ( $X_3$ ).

### 4.1 Regresi linier berganda

Setelah data dianalisis, data terbukti memenuhi asumsi normalitas yang ditunjukkan dari nilai  $p$ -value dari uji Anderson Darling yaitu  $p$ -value = 0,737 > 0,05. Selanjutnya, data tidak terdapat autokorelasi ditunjukkan dari uji Durbin-Watson bahwa  $d = 2,2429 > d_u = 1,676$  artinya tidak terdapat autokorelasi positif dan  $(4 - d) = 1,7571 > d_u = 1,676$  artinya tidak terdapat autokorelasi negatif. Selanjutnya, data tidak mengandung heteroskedastisitas yang ditunjukkan dari uji Glejser. Selanjutnya, nilai VIF antar variabel bebas yaitu  $VIF_1 = 708,8429$ ;  $VIF_2 = 564,3434$  dan  $VIF_3 = 104,6060$  melebihi nilai 10. Artinya data pada teladan ini terdeteksi adanya

<sup>1</sup>Alumni Jurusan Matematika FMIPA Universitas Bengkulu

<sup>2,3</sup>Staf Pengajar Jurusan Matematika FMIPA Universitas Bengkulu

multikolinieritas. Akibat dari multikolinieritas, secara individu variabel bebas tidak berpengaruh terhadap variabel respon, tetapi nilai koefisien determinasi ( $R^2$ ) bernilai 80,14%. Sehingga diperlukan cara dalam mengatasi masalah multikolinieritas yaitu salah satunya digunakan metode *Latent Root Regression* (LRR).

#### 4.2 Pembakuan data

$Z_y$  hasil pembakuan data variabel  $Y$  yang dihitung dari persamaan (3.3) dan  $Z_j$  hasil pembakuan data variabel  $X_j$  yang dihitung dari persamaan (3.4) untuk  $i = 1, 2, \dots, 20$  dan  $j = 1, 2, 3$ .

#### 4.3 Perhitungan matriks korelasi gabungan

Terlebih dahulu menghitung matriks gabungannya yaitu

$$\begin{aligned} \mathbf{Z}^*_{20 \times 4} &= [\mathbf{Z}_y, \mathbf{Z}]_{20 \times 4} \\ &= \begin{bmatrix} -0,3727 & -0,2651 & -0,3537 & 0,0931 \\ 0,1170 & -0,0276 & -0,0600 & 0,0365 \\ \vdots & \vdots & \vdots & \vdots \\ 0,0407 & -0,0048 & -0,0075 & -0,0075 \end{bmatrix}_{20 \times 4} \end{aligned}$$

Selanjutnya, menghitung matriks korelasi gabungan yakni diperoleh

$$\mathbf{Z}^{*'} \mathbf{Z}^* = \begin{bmatrix} 1 & 0,8433 & 0,8781 & 0,1424 \\ 0,8433 & 1 & 0,9238 & 0,4578 \\ 0,8781 & 0,9238 & 1 & 0,0847 \\ 0,1424 & 0,4578 & 0,0847 & 1 \end{bmatrix}_{4 \times 4}$$

#### 4.4 Akar laten dan vektor laten

Berdasarkan hasil dari perhitungan matriks korelasi gabungan maka dapat dicari akar laten  $\lambda_j$  dan vektor laten  $\Gamma_j$  dengan  $j = 0, 1, 2, 3$  yang bersesuaian dengan  $\lambda_j$  dibentuk dari matriks korelasi gabungan  $\mathbf{Z}^{*'} \mathbf{Z}^*$ . Hasil nilai akar laten dan vektor laten yang bersesuaian disajikan pada Tabel 1 berikut:

Tabel 1. Nilai akar laten dan vektor laten

akar Laten $\lambda_j$	$Z_y$ $\gamma_{0j}$	$Z_1$ $\gamma_{1j}$	$Z_2$ $\gamma_{2j}$	$Z_3$ $\gamma_{3j}$
0,0007	0,0178	-0,7248	0,6310	0,2760
0,1484	0,8079	-0,3479	-0,4654	0,0981
1,0013	-0,2151	0,1186	-0,2655	0,9323
2,8497	0,5483	0,5827	0,5611	0,2121

#### 4.5 Pembentukan komponen utama

Berdasarkan hasil akar laten dan vektor laten di atas, dapat diketahui bahwa terdapat akar laten yang nilainya kurang dari 0,05 yaitu  $\lambda_0 = 0,0007 \leq 0,05$ . Selanjutnya, memeriksa vektor laten yang bersesuaian dengan  $\lambda_0$ . Untuk  $\gamma_{00}$  nilainya 0,0178, artinya nilai  $|\gamma_{00} = 0,0178| < 0,10$ . Oleh karena itu, akar laten dan vektor laten yang bersesuaian tersebut harus dihapus. Sehingga komponen utama yang digunakan yaitu

<sup>1</sup>Alumni Jurusan Matematika FMIPA Universitas Bengkulu

<sup>2,3</sup>Staf Pengajar Jurusan Matematika FMIPA Universitas Bengkulu

$$\begin{aligned}
PC_1 &= 0,8079Z_y - 0,3479Z_1 - 0,4654Z_2 + 0,0981Z_3 \\
PC_2 &= -0,2151Z_y + 0,1186Z_1 - 0,2655Z_2 + 0,9323Z_3 \\
PC_3 &= 0,5483Z_y + 0,5827Z_1 + 0,5611Z_2 + 0,2121Z_3
\end{aligned}$$

Komponen utama tersebut diregresikan dengan variabel respon untuk melihat apakah masalah multikolinieritas telah teratasi. Proses analisis regresi tersebut dinamakan analisis regresi akar laten (*Latent Root Regression*). Berdasarkan analisis regresi akar laten diperoleh model regresi yang terbentuk yaitu:

$$Y = 20,1950 + 17,9823PC_1 - 4,7873PC_2 + 12,2045PC_3$$

Berdasarkan hasil analisis regresi akar laten diperoleh nilai *VIF* masing-masing dari komponen utama sebesar 1,000 dan nilai korelasi antar  $PC_1$  dan  $PC_2$  bernilai  $1,3797 \times 10^{-9}$ ,  $PC_1$  dan  $PC_3$  bernilai  $1,3871 \times 10^{-8}$  dan  $PC_2$  dan  $PC_3$  bernilai  $-9,7716 \times 10^{-9}$ . Hal ini menunjukkan bahwa masalah multikolinieritas dapat diatasi secara tuntas.

#### 4.6 Koefisien regresi kuadrat terkecil

Hasil koefisien regresi kuadrat terkecil dapat dilihat pada Tabel 2 berikut:

Tabel 2. Koefisien regresi kuadrat terkecil

Koefisien	<i>j</i>			
	0	1	2	3
<b>b*</b>	20,1950	8,8411	11,5862	-1,8337
<b>b</b>	-12,8201	0,4038	0,5078	-0,1153

dengan  $b^*$  merupakan koefisien regresi kuadrat terkecil termodifikasi untuk variabel yang telah dibakukan, sedangkan  $b$  merupakan koefisien regresi kuadrat terkecil untuk variabel awal. Sehingga model regresi untuk data awal diperoleh sebagai berikut

$$Y = -12,8201 + 0,4038X_1 + 0,5078X_2 - 0,1153X_3$$

Model regresi di atas dapat diinterpretasikan bahwa pada saat semua variabel bebas konstan, maka variabel persentase lemak tubuh akan bernilai  $-12,8201$ . Variabel persentase lemak tubuh akan meningkat sebesar 0,4038 setiap kenaikan  $X_1$  satu mm selama  $X_2$  dan  $X_3$  bernilai konstan. Variabel persentase lemak tubuh akan meningkat sebesar 0,5078 setiap kenaikan  $X_2$  satu mm selama  $X_1$  dan  $X_3$  bernilai konstan. Variabel persentase lemak tubuh akan berkurang sebesar 0,1153 setiap kenaikan  $X_3$  satu mm selama  $X_1$  dan  $X_2$  bernilai konstan.

Selanjutnya, untuk melihat keakuratan model yang diperoleh, dapat dilihat dari koefisien determinasi yaitu bernilai 100% artinya model dapat dikatakan baik atau akurat. Nilai koefisien determinasi sebesar 100% bermakna bahwa sebesar 100% variansi dari besarnya nilai persentase lemak tubuh ( $Y$ ) dapat dijelaskan oleh *triceps skinfold thickness* ( $X_1$ ), *thigh circumference* ( $X_2$ ) dan *mid arm circumference* ( $X_3$ ) dalam hubungan linier.

#### 4.7 Metode eliminasi backward

Selanjutnya, dilakukan proses eliminasi *backward* (langkah mundur) untuk mengeluarkan variabel bebas yang tidak berpengaruh terhadap model. Hasil uji- $F$  yang dihasilkan dengan penghapusan variabel  $X_r$  dengan  $r = 1, 2, 3$  yaitu

<sup>1</sup>Alumni Jurusan Matematika FMIPA Universitas Bengkulu

<sup>2,3</sup>Staf Pengajar Jurusan Matematika FMIPA Universitas Bengkulu

Tabel 3. Pengujian eliminasi *backward*

Variabel	$n - k - 1$	$F_{tabel}$	$F_{hitung}$
$X_1$	16	0,0041	65,7546
$X_2$			64,4421
$X_3$			16,5384

Berdasarkan Tabel 3 di atas, untuk menguji variabel  $X_1$  dihapus dari model terlihat dari nilai  $F_{hitung} = 65,7546 > 0,0041$  artinya variabel  $X_1$  tetap dipertahankan dalam model. Selanjutnya, untuk menguji variabel  $X_2$  dihapus dari model terlihat nilai  $F_{hitung} = 64,4421 > 0,0041$  artinya variabel  $X_2$  tetap dipertahankan dalam model. Selanjutnya, untuk menguji variabel  $X_3$  dihapus dari model terlihat nilai  $F_{hitung} = 16,5384 > 0,0041$  artinya variabel  $X_3$  tetap dipertahankan dalam model. Sehingga dapat dikatakan bahwa tidak ada variabel bebas yang dihapus atau dikeluarkan dari model.

## 5. KESIMPULAN DAN SARAN

### 5.1 Kesimpulan

Berdasarkan hasil penelitian maka dapat disimpulkan bahwa:

1. Multikolinieritas dapat dideteksi dengan cara melihat nilai VIF (*Variance Inflation Factor*) yang melebihi nilai 10.
2. Prosedur metode LRR (*Latent Root Regression*) untuk mengatasi adanya multikolinieritas yaitu:
  - Membakukan data (variabel respon  $Y$  dan variabel bebas  $X$ )
  - Membuat matriks gabungan dari variabel yang telah dibakukan, dilambangkan dengan  $Z^* = [Z_y : Z]$
  - Menghitung matriks korelasi gabungan dari matriks  $Z^*$
  - Menghitung akar laten dan vektor laten berdasarkan matriks korelasi gabungan
  - Memilih komponen utama yang digunakan dengan membuang komponen utama yang mempunyai akar laten  $\lambda_j \leq 0,05$  dan elemen pertama dari vektor laten  $|\gamma_{0j}| < 0,10$
  - Berdasarkan langkah 5, komponen utama yang telah ditentukan tersebut diregresikan dengan variabel respon
  - Menghitung nilai korelasi dan nilai VIF antar variabel komponen utama untuk mendeteksi apakah masalah multikolinieritas sudah teratasi
  - Menghitung pendugaan koefisien regresi pada data yang telah dibakukan
  - Menghitung endugaan koefisien regresi pada variabel awal
  - Menerapkan metode eliminasi *backward*.

### 5.2 Saran

Dalam penulisan skripsi ini, penulis hanya membahas pelanggaran asumsi multikolinieritas yaitu mengatasinya dengan menggunakan metode *Latent Root Regression* (LRR). Untuk penelitian selanjutnya, dapat dibahas mengenai perbandingan metode regresi komponen utama dan regresi akar laten dalam mengatasi masalah multikolinieritas.

<sup>1</sup>Alumni Jurusan Matematika FMIPA Universitas Bengkulu

<sup>2,3</sup>Staf Pengajar Jurusan Matematika FMIPA Universitas Bengkulu

## DAFTAR PUSTAKA

- Bilfarsah, A, 2005, Efektifitas Metode Aditif Spline Kuadrat Terkecil Parsial dalam Pendugaan Model Regresi, *Makara, Sains*, Vol. 9, No.1, Halm: 28-33.
- Draper, N. R dan H. Smith, 1992, *Analisis Regresi Terapan*, Edisi Kedua, Diterjemahkan oleh Bambang Sumantri, Jurusan Statistika FMIPA IPB, Bogor.
- Gujarati, D. N, 2003, *Basic Econometrics, Fourth Edition*, McGraw-Hill Higher Education, New York.
- Hoerl, A. E dan R. W. Kennard, 1970, Ridge Regression: Biased Estimation for Non Orthogonal Problem. *Technometrics*, Vol. 12, No. 1, Halm: 55-67.
- Johnson, R. A dan D. W, Wichern, 2002, *Applied Multivariate Statistical Analysis, Fifth Edition*, Pearson Education International.
- Kutner, M. H., C. J, Nachtsheim., J. Neter dan W. Li, 2005, *Applied Linear Statistical Model, Fifth Edition*, McGraw-Hill, New York.
- Marcus, G. L., H. J, Wattimanela dan Y. A, Lesnussa, 2012, Analisis Regresi Komponen Utama untuk Mengatasi Masalah Multikolinieritas dalam Analisis Regresi Linier Berganda, *Jurnal Barekeng*, Vol. 7, No.1, Halm: 31-40.
- Myers, R. H dan J. S, Milton, 1991, *A First Course In The Theory of Linier Statistical Models*, PWS-KENT Publihing Company, Boston.
- Santosa, R. G, 2009, *Aljabar Linier Dasar*, ANDI, Yogyakarta.
- Sembiring, R. K, 2003, *Analisis Regresi*, ITB, Bandung.
- Sharma, S dan W. L, James, 1986, Latent Root Regression: An Alternate Procedure for Estimating Parameters in the Presence of Multicollinearity, *Journal of Marketing Research*, Vol. 18, No. 2, PP: 154-161.
- Soelistyo, 2000, *Dasar-dasar Ekonometrika*, Edisi pertama, BPFE, Yogyakarta.
- Webster, J. T., R. F. Gunst dan R. L. Mason. 1974. Latent Root Regression Analysis. *Technometrics*, Vo. 16, No.4, PP: 513-522.

<sup>1</sup>Alumni Jurusan Matematika FMIPA Universitas Bengkulu

<sup>2,3</sup>Staf Pengajar Jurusan Matematika FMIPA Universitas Bengkulu

# METODE STANDARD ERROR NEWKEY WEST UNTUK MENGATASI HETEROSKEDASTISITAS DAN AUTOKORELASI PADA ANALISIS REGRESI DATA PANEL

Bella Haryati<sup>1</sup>, Sigit Nugroho<sup>2</sup> dan Etis Sunandi<sup>3</sup>

<sup>1</sup>Alumni Jurusan Matematika, FMIPA, Universitas Bengkulu

<sup>2,3</sup>Staf Pengajar Jurusan Matematika, FMIPA, Universitas Bengkulu

e-mail: [bellaharyati.bh@gmail.com](mailto:bellaharyati.bh@gmail.com), [sigit.nugroho.1960@gmail.com](mailto:sigit.nugroho.1960@gmail.com),

[etiss18@gmail.com](mailto:etiss18@gmail.com)

## ABSTRAK

Penelitian ini bertujuan untuk mengatasi heteroskedastisitas dan autokorelasi pada analisis regresi data panel dengan menggunakan metode *standard error Newey West*. Data yang digunakan adalah data sekunder, yaitu rasio ketersediaan beras, ketersediaan beras, luas panen, produktivitas lahan, jumlah konsumsi beras dan harga beras dari tahun 2009-2014. Data tersebut melanggar asumsi non autokorelasi dan homoskedastisitas. Dari hasil analisis didapatkan informasi bahwa ketersediaan beras, luas panen, produktivitas lahan dan jumlah konsumsi beras berpengaruh nyata terhadap rasio ketersediaan beras. Selain itu, *standard error* yang didapatkan dari metode *Newey West*, dengan nilai  $g = 1$ , dan  $g = 4$  selalu memiliki nilai lebih besar dari *standard error* awal.

**Kata kunci:** Regresi data panel, *standard error Newey West*, heteroskedastisitas, autokorelasi

## 1. PENDAHULUAN

Di Indonesia pangan sering diidentikkan dengan beras karena jenis pangan ini merupakan makanan pokok utama bagi masyarakatnya. Tingginya konsumsi beras tergambar dari besarnya alokasi pengeluaran untuk beras dalam struktur pengeluaran keluarga. Di Bengkulu *World Bank* memperkirakan 34% dari 70% pengeluaran keluarga miskin dialokasikan untuk membeli beras sebagai makanan pokok (Supardi, 2013).

Pada kehidupan sehari-hari sering dijumpai perkiraan ketahanan pangan yang hanya memasukkan salah satu unsur runtun waktu atau data silang. Data runtun waktu adalah data yang diamati pada satu unit berdasarkan runtun waktu tertentu. Dalam analisis runtun waktu, lebih besar kemungkinan terjadi autokorelasi positif, karena variabel yang dianalisis biasanya mengandung kecenderungan meningkat.

Autokorelasi terjadi karena beberapa sebab. Salah satu penyebabnya menurut Gujarati (2004), adalah data runtun waktu. Sedangkan data data silang adalah data yang terdiri dari beberapa objek yang dikumpulkan pada satu waktu tertentu. Data data silang cenderung untuk bersifat heteroskedastisitas karena pengamatan dilakukan pada individu yang berbeda pada saat yang sama. Gabungan dari kedua data ini dinamakan data panel. Salah satu keuntungan yang diperoleh dengan menggunakan data panel adalah mampu menyediakan data yang lebih banyak sehingga akan menghasilkan derajat kebebasan yang lebih besar (Apriliawan, Tarno, dan Yasin, 2013).

Analisis regresi merupakan sebuah alat statistik yang digunakan untuk mengetahui pengaruh antara dua atau lebih variabel sehingga sebuah respon dapat diprediksi. Analisis regresi memiliki

<sup>1</sup>Alumni Jurusan Matematika FMIPA Universitas Bengkulu

<sup>2,3</sup>Staf Pengajar Jurusan Matematika FMIPA Universitas Bengkulu

beberapa asumsi klasik yang harus dipenuhi untuk mendapatkan estimasi parameter pada model yang bersifat *BLUE* (*Best Linear Unbiased Estimator*). Namun dalam penerapannya sering sekali terjadi pelanggaran asumsi. Asumsi yang sering tidak terpenuhi seperti asumsi homoskedastisitas dan non-autokorelasi sehingga estimasi parameter tidak lagi bersifat *BLUE*. Selain itu, Standard error akan *underestimate* dari *standard error* yang sebenarnya (Kutner, 2004).

*Standard error* Newey West dapat mengoreksi *standard error* yang didapatkan sehingga *standard error* tidak akan terlalu jauh dari parameter (*underestimate*). Metode *standard error* Newey West yang dikembangkan pada tahun 1987 merupakan perluasan dari metode *standard error* White. Metode *standard error* White hanya kekar terhadap heteroskedastisitas, sedangkan metode *standard error* Newey West kekar terhadap autokorelasi dan heteroskedastisitas (Wooldrige, 2009).

Pada penelitian sebelumnya Silalahi, Sitepu, dan Tarigan (2014) telah menerapkan analisis regresi data panel dengan studi kasus ketahanan pangan Provinsi Sumatera Utara tahun 2007-2011 dengan mendeteksi masalah heteroskedastisitas dan autokorelasi menggunakan uji Park dan uji Durbin Watson. Pada penelitian ini, penulis akan menggunakan *standard error* Newey West untuk mendeteksi masalah heteroskedastisitas dan autokorelasi.

## 2. LANDASAN TEORI

### 2.1. Regresi Data Panel

Regresi data panel adalah analisis regresi yang digunakan pada data panel. Menurut Hsiao (2014) data panel merupakan data gabungan dari dua tipe data yaitu data runtun waktu dan data silang. Data runtun waktu adalah data dari pengamatan yang dilakukan pada satu individu pada runtun waktu yang dapat

berubah-ubah seiring berjalannya waktu dengan periode waktu yang sama. Sedangkan data silang adalah data dari pengamatan yang dilakukan pada individu yang berbeda pada waktu yang sama. Bentuk umum regresi data panel:

$$Y_{it} = \alpha_{it} + \sum_{j=1}^P \beta_j X_{jit} + u_{it} \quad (1)$$

dengan:

$Y_{it}$  : variabel terikat untuk individu ke- $i$  pada waktu ke- $t$

$\alpha_{it}$  : intersep untuk individu ke- $i$  pada waktu ke- $t$

$\beta_j$  : parameter untuk variabel ke- $j$

$X_{jit}$  : variabel bebas ke- $j$  untuk individu ke- $i$  pada waktu ke- $t$

$u_{it}$  : *error* regresi untuk individu ke- $i$  pada waktu ke- $t$

$i$  : unit data silang sebanyak  $N$

$t$  : unit runtun waktu sebanyak  $T$

karena data bersifat panel maka *error* dari regresi ini memiliki komponen yang umum dan spesifik. Karakter ini secara matematis ditunjukkan oleh persamaan (2):

$$u_{it} = e + v_i + w_t \quad (2)$$

dimana  $e$  adalah komponen *error* yang bersifat umum,  $v_i$  adalah komponen yang spesifik data silang dan  $w_t$  adalah komponen yang spesifik runtun waktu (Ariefianto, 2012).

### 2.2. Penduga Parameter Model Regresi Data Panel

Dalam menduga parameter model regresi pada data panel ada tiga macam pendekatan yaitu Model Efek Gabungan, Model Efek Tetap dan Model Efek Acak.

#### 2.2.1. Model Efek Gabungan

Model Efek Gabungan adalah metode regresi yang menduga data panel dengan metode *Ordinary Least Square* (OLS). Metode ini tidak memperhatikan dimensi individu maupun waktu sehingga diasumsikan bahwa perilaku antar individu sama dalam berbagai kurun

<sup>1</sup>Alumni Jurusan Matematika FMIPA Universitas Bengkulu

<sup>2,3</sup>Staf Pengajar Jurusan Matematika FMIPA Universitas Bengkulu

waktu. Persamaan metode ini dapat ditulis sebagai berikut:

$$Y_{it} = \alpha + \sum_{j=1}^P \beta_j X_{jit} + u_{it} \quad (3)$$

dengan:

- $Y_{it}$  : variabel terikat untuk individu ke- $i$  pada waktu ke- $t$
- $\alpha$  : intersep
- $\beta_j$  : parameter untuk variabel ke- $j$
- $X_{jit}$  : variabel bebas ke- $j$  untuk individu ke- $i$  pada waktu ke- $t$
- $u_{it}$  : komponen *error* untuk individu ke- $i$  pada waktu ke- $t$
- $i$  : unit data silang sebanyak  $N$
- $t$  : unit runtun waktu sebanyak  $T$

### 2.2.2. Model Efek Tetap

Model Efek Tetap adalah metode regresi yang menduga data panel dengan menambahkan variabel *dummy*. Teknik ini dinamakan *Least Square Dummy Variabel* (LSDV). Teknik ini diterapkan untuk efek tiap individu. Model Efek Tetap mengasumsikan bahwa terdapat efek yang berbeda antar individu. Setiap individu merupakan parameter yang tidak diketahui dan akan diduga dengan menggunakan teknik variabel *dummy* yang dapat ditulis sebagai berikut:

$$Y_{it} = \alpha_i + \sum_{j=1}^P \beta_j X_{jit} + u_{it} \quad (4)$$

dengan:

- $Y_{it}$  : variabel terikat untuk individu ke- $i$  pada waktu ke- $t$
- $\alpha_i$  : intersep
- $\beta_j$  : parameter untuk variabel ke- $j$
- $X_{jit}$  : variabel bebas ke- $j$  untuk individu ke- $i$  pada waktu ke- $t$
- $i$  : unit data silang sebanyak  $N$
- $t$  : unit runtun waktu sebanyak  $T$
- $u_{it}$  : komponen *error* untuk individu ke- $i$  pada waktu ke- $t$

Dalam menduga parameter untuk data panel, Gujarati (2004) menyatakan

Model Efek Tetap memiliki beberapa kelemahan yakni:

- a. Masalah kekurangan derajat kebebasan akibat jumlah contoh yang terbatas.
- b. Multikolinieritas yang diakibatkan oleh banyaknya variabel *dummy* yang diduga.
- c. Keterbatasan kemampuan penduga, terutama jika terdapat variabel yang bersifat tidak berubah berdasarkan waktu.
- d. Kemungkinan korelasi di antara komponen residual spesifik (data silang dan runtun waktu) (Ariefianto, 2012).

### 2.2.3. Model Efek Acak

Model Efek Acak adalah Model regresi yang menduga parameter regresi data panel dengan menghitung *error* dari model regresi dengan metode *Generalized Least Square* (GLS). Persamaan Model Efek Acak dapat ditulis sebagai berikut:

$$Y_{it} = \alpha_i + \sum_{j=1}^P \beta_j X_{jit} + u_{it} \quad (5)$$

dengan:

- $Y_{it}$  : variabel terikat untuk individu ke- $i$  pada waktu ke- $t$
- $\alpha_i$  : intersep,  $\alpha_i = \alpha_0 + \varepsilon_i$
- $\beta_j$  : parameter untuk variabel ke- $j$
- $X_{jit}$  : variabel bebas ke- $j$  untuk individu ke- $i$  pada waktu ke- $t$
- $u_{it}$  : komponen *error* untuk individu ke- $i$  pada waktu ke- $t$
- $i$  : unit data silang sebanyak  $N$
- $t$  : unit runtun waktu sebanyak  $T$

### 2.3. Pemilihan Model

Langkah-langkah yang dilakukan dalam mendapatkan model yang tepat adalah pertama dilakukan uji Chow. Jika dalam uji Chow terbukti tidak ada efek individu maka digunakan Model Efek Gabungan. Namun jika dalam uji Chow terbukti ada efek individu maka selanjutnya dilakukan uji Hausman untuk

<sup>1</sup>Alumni Jurusan Matematika FMIPA Universitas Bengkulu

<sup>2,3</sup>Staf Pengajar Jurusan Matematika FMIPA Universitas Bengkulu

menentukan antara Model Efek Tetap dan Model Efek Acak.

### 2.3.1. Uji Chow

Uji Chow digunakan untuk memilih Model Efek Gabungan atau Model Efek Tetap dengan hipotesis sebagai berikut:

1. Merumuskan hipotesis, yaitu:  
 $H_0$ : Model Efek Gabungan  
 $H_1$ : Model Efek Tetap
2. Tingkat signifikansi  $\alpha = 5\%$
3. Statistik Uji

$$F = \frac{[R_{LSDV}^2 - R_{pooled}^2]/(N - 1)}{(1 - R_{LSDV}^2)/(NT - N - K)} \quad (6)$$

dengan:

- $R_{pooled}^2$  : *R - square* untuk Model Efek Gabungan  
 $R_{LSDV}^2$  : *R - square* untuk Model Efek Tetap  
 $N$  : jumlah unit data silang  
 $T$  : jumlah unit runtun waktu  
 $K$  : jumlah variabel bebas

4. Kriteria Penolakan

Tolak  $H_0$  jika  $F_{hitung} >$

$F_{(N-1, NT-N-K, \alpha)}$  atau  $p - value < \alpha$

5. Kesimpulan

Jika  $H_0$  ditolak artinya model penduga yang tepat digunakan adalah metode pendekatan Model Efek Tetap.

### 2.3.2. Uji Hausman

Uji Hausman merupakan pengujian untuk memilih Model Efek Tetap atau Model Efek Acak. Uji ini memiliki hipotesis sebagai berikut:

1. Merumuskan hipotesis, yaitu:  
 $H_0$ : Model Efek Tetap  
 $H_1$ : Model Efek Acak
2. Tingkat signifikansi  $\alpha = 5\%$
3. Statistik Uji

$$W = (b - \hat{\beta})[var(b - \hat{\beta})]^{-1}(b - \hat{\beta}) \quad (7)$$

dengan:

- $b$  : koefisien efek acak  
 $\hat{\beta}$  : koefisien efek tetap

4. Kriteria Penolakan

Tolak  $H_0$  jika  $W > \chi^2_{(K, \alpha)}$  atau  $p - value < \alpha$

5. Kesimpulan

Jika  $H_0$  ditolak artinya model penduga yang tepat digunakan adalah metode pendekatan Model Efek Acak.

### 2.4. Standard Error

*Standard error* (SE) atau kesalahan baku adalah indeks yang menggambarkan sebaran rata-rata contoh terhadap rata-rata keseluruhan kemungkinan contoh (rata-rata populasi). Prinsip OLS adalah meminimalkan *error*. Oleh karena itu, ketepatan dari nilai dugaan sangat ditentukan oleh *standard error* dari masing-masing penduga. Menurut Ariefianto (2012), *standard error* dapat dinyatakan sebagai berikut:

$$SE(\hat{\beta}_k) = \sqrt{Var(\hat{\beta}_k)} \quad (8)$$

dengan nilai  $Var(\hat{\beta}_k)$  dapat diperoleh dari formula sebagai berikut (Rachmawati dan Sumarminingsih, 2014):

$$Var(\hat{\beta}_k) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (9)$$

dengan:

- $Se$  : nilai *standard error*  
 $Var(\hat{\beta}_k)$  : ragam parameter ke- $k$

### 2.5. Uji Asumsi Klasik

Asumsi yang harus dipenuhi adalah kenormalan, tidak ada multikolinieritas, homoskedastisitas, dan non autokorelasi. Asumsi normalitas digunakan untuk mengetahui apakah *error* data berdistribusi normal atau tidak. Dalam uji ini diharapkan asumsi normalitas terpenuhi. Sedangkan asumsi multikolinieritas adalah asumsi untuk mengetahui ada atau tidaknya korelasi yang signifikan antara variabel-variabel bebas dalam suatu model regresi linier berganda. Dalam uji ini diharapkan tidak ada multikolinieritas. Untuk asumsi

<sup>1</sup>Alumni Jurusan Matematika FMIPA Universitas Bengkulu

<sup>2,3</sup>Staf Pengajar Jurusan Matematika FMIPA Universitas Bengkulu

homoskedastisitas dan non autokorelasi akan dibahas pada bagian selanjutnya.

### 2.5.1. Uji Heteroskedastisitas

Asumsi penting yang harus dipenuhi ialah ragam dari *error* tidak berubah, dengan berubahnya satu atau lebih variabel bebas yang sering disebut homoskedastisitas. Namun jika asumsi ini tidak terpenuhi maka disebut heteroskedastisitas. Heteroskedastisitas terjadi apabila variabel gangguan tidak mempunyai varian yang sama untuk semua pengamatan (Gujarati, 2004). Mendeteksi heteroskedastisitas dengan menggunakan uji White adalah sebagai berikut:

1. Merumuskan hipotesis, yaitu:  
 $H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_t^2 = \sigma^2 = 0$   
 $H_1: \text{paling tidak ada satu } \sigma_i^2 \text{ dimana } \sigma_i^2 \neq \sigma^2, i = 1, 2, \dots, n.$
2. Menentukan taraf nyata ( $\alpha$ ) sebesar 5%
3. Statistik Uji yang digunakan

$$White = nR^2 \quad (11)$$

dengan :

$n$  : jumlah contoh  
 $R^2$  : nilai *R - square*

4. Kriteria Penolakan  
 $White < \chi^2$ , berarti  $H_0$  diterima dan  $H_1$  ditolak. Artinya *error* bersifat homoskedastis  
 $White > \chi^2$ , berarti  $H_0$  ditolak dan  $H_1$  diterima. Artinya *error* bersifat heteroskedastis

### 2.5.2. Uji Autokorelasi

Menurut Gujarati (2004), autokorelasi merupakan korelasi diantara anggota seri pengamatan yang disusun menurut runtun waktu atau menurut tempat/ruang (dalam data silang). Autokorelasi atau sering juga disebut korelasi serial merupakan suatu bentuk pelanggaran terhadap asumsi klasik.

Gejala autokorelasi dalam penelitian ini dideteksi menggunakan metode Durbin-Watson (DW). Penentuan ada atau tidaknya autokorelasi menurut metode

Durbin-Watson (DW) dapat dijelaskan pada tabel berikut:

Tabel 1. Uji statistik d Durbin-Watson (DW)

Nilai statistik d	Hasil
$0 \leq d \leq d_L$	$H_0$ ditolak: ada autokorelasi positif
$d_L \leq d \leq d_U$	Daerah keragu-raguan: tidak ada keputusan
$2 \leq d \leq 4 - d_U$	Gagal menolak $H_0$ tidak ada autokorelasi
$4 - d_U \leq d \leq 4 - d_L$	Daerah keragu-raguan: tidak ada keputusan
$4 - d_L \leq d \leq 4$	$H_0$ ditolak: ada autokorelasi negatif

dimana  $d_L$  dan  $d_U$  adalah batas bawah dan batas atas nilai kritis yang dapat dicari dari tabel Durbin Watson berdasarkan  $k$  (jumlah variabel bebas) dan  $n$  (jumlah contoh) yang relevan. Berikut uji autokorelasi menggunakan uji Durbin Watson dengan hipotesis sebagai berikut (Gujarati, 2004):

1. Merumuskan hipotesis, yaitu:  
 $H_0: \rho_1 = \rho_2 = \dots = \rho_p = 0$  (tidak ada autokorelasi)  
 $H_1: \text{paling tidak ada satu } \rho_p \text{ dimana } \rho_p \neq 0, i = 1, 2, \dots, p$
2. Menentukan taraf nyata ( $\alpha$ ) sebesar 5%
3. Statistik Uji

$$d = \frac{\sum_{t=2}^n (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^n \hat{u}_t^2} \quad (13)$$

dengan:

$d$  : nilai durbin watson  
 $\hat{u}_t$  : *error* regresi untuk waktu ke- $t$

4. Kriteria Penolakan  
 Dengan melihat nilai  $DW$  terletak diantara  $2 < DW < 4 - d_U$  untuk menentukan ada tidaknya autokorelasi. Dimana  $d_L$  dan  $d_U$  adalah batas bawah dan batas atas.

### 2.6. Metode Standard Error Newey West

Dampak dari adanya autokorelasi adalah *standard error* parameter menjadi bias. Dengan demikian salah satu cara

<sup>1</sup>Alumni Jurusan Matematika FMIPA Universitas Bengkulu

<sup>2,3</sup>Staf Pengajar Jurusan Matematika FMIPA Universitas Bengkulu

untuk mengoreksi kondisi ini, menurut Ariefianto (2012) adalah dengan membuat formulasi *standard error* yang tidak bias.

Wooldridge (2009) mengatakan bahwa metode *standard error* Newey West yang dikembangkan pada tahun 1987 merupakan perluasan dari metode *standard error* White. Metode *standard error* White hanya kekar terhadap heteroskedastisitas, sedangkan metode *standard error* Newey West kekar terhadap autokorelasi dan heteroskedastisitas. Rumus *standard error* Newey West adalah sebagai berikut:

$$se_{newey-west}(\hat{\beta}_p) = \left( \frac{se(\hat{\beta}_p)}{\hat{\sigma}^2} \right)^2 \sqrt{\hat{v}} \quad (14)$$

dengan:

$se(\hat{\beta}_p)$  : *standard error* parameter  $p$  dari regresi awal  
 $\hat{\sigma}^2$  : estimator ragam model regresi awal

Nilai  $\hat{v}$  dapat dihitung dengan menggunakan rumus

$$\hat{v} = \sum_{t=1}^n \hat{a}_t^2 + 2 \sum_{n=1}^g \left[ 1 - \frac{h}{g+1} \right] \left( \sum_{t=h+1}^n \hat{a}_t \hat{a}_{t-h} \right) \quad (15)$$

dengan nilai  $\hat{a}_t$  diperoleh dengan langkah sebagai berikut:

- $y_t = \beta_0 + \beta_1 x_{1t} + \dots + \beta_k x_{kt} + u_t$  dengan OLS untuk memperoleh  $se(\beta_1)$ ,  $\sigma$ , dan *error*  $u_t$ .
- Hitung *error*  $r_t$  dari *auxiliary* regresi (regresi tambahan):

$$x_{1t} = \delta_0 + \delta_2 x_{t2} + \dots + \delta_k x_{tk} + r_t$$

$$\hat{a}_t = \hat{r}_t \hat{u}_t, \quad t = 1, 2, \dots, n$$

Nilai  $g$  dapat disesuaikan dengan data, data tahunan dipilih  $g = 1$ . Newey West menyarankan mengambil  $g$  menjadi bagian integer dari  $4 \left( \frac{n}{100} \right)^{2/9}$ . Nilai  $h$  diperoleh dengan rumus  $h = \exp(g)$ .

## 2.7. Uji Signifikansi

Data yang telah diolah, perlu dilakukan pengujian terhadap model yang dihasilkan agar dapat memberikan

gambaran yang jelas atas apa yang ingin diinformasikan dari model tersebut.

### 2.7.1. Uji Simultan (Uji F)

Uji simultan digunakan untuk melihat pengaruh variabel bebas secara bersama-sama terhadap variabel terikat. Langkah-langkah yang harus dilakukan untuk uji F sebagai berikut:

- Merumuskan hipotesis, yaitu:

$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$  artinya variabel bebas secara simultan tidak berpengaruh terhadap variabel terikat

$H_1: \beta_i \neq 0$  artinya variabel bebas secara simultan berpengaruh terhadap variabel terikat.  $\forall i = 1, 2, \dots, k$

- Menentukan taraf nyata ( $\alpha$ : 5%) dan derajat kebebasan ( $db$ ) untuk mengetahui nilai  $F_{tabel} = F_{(\alpha; N_1, N_2)}$ . Dimana  $n$  adalah jumlah contoh dan  $k$  adalah jumlah variabel bebas.
- Mencari nilai  $F_{hitung}$  dengan rumus dinyatakan sebagai berikut:

$$F_{hitung} = \frac{(R^2)/(N-1)}{(1-R^2)/(NT-N-K)} \quad (18)$$

dengan:

$R^2$  : koefisien determinasi  
 $T$  : jumlah individu / data silang  
 $N$  : jumlah runtun waktu  
 $K$  : jumlah variabel bebas

- Kriteria Penolakan

$F_{hitung} < F_{(\alpha; N_1, N_2)}$ , berarti  $H_0$  diterima dan  $H_1$  ditolak. Artinya variabel secara bersama tidak berpengaruh terhadap variabel terikat.

$F_{hitung} > F_{(\alpha; N_1, N_2)}$ , berarti  $H_0$  ditolak dan  $H_1$  diterima. Artinya variabel secara bersama berpengaruh terhadap variabel terikat.

### 2.7.2 Uji Parsial (Uji t)

Uji parsial atau uji t dipakai untuk melihat signifikansi pengaruh variabel bebas secara individu terhadap variabel terikat dengan menganggap variabel lain

<sup>1</sup>Alumni Jurusan Matematika FMIPA Universitas Bengkulu

<sup>2,3</sup>Staf Pengajar Jurusan Matematika FMIPA Universitas Bengkulu

bersifat konstan. Uji ini dilakukan dengan membandingkan  $t_{hitung}$  dengan  $t_{tabel}$ . Adapun langkah-langkah yang harus dilakukan untuk uji T adalah sebagai berikut:

1. Merumuskan hipotesis, yaitu:  
 $H_0: \beta_i = 0$  artinya secara individu variabel ke- $i$  tidak berpengaruh terhadap variabel terikat.  
 $H_1: \beta_i \neq 0$  artinya variabel ke- $i$  secara individu berpengaruh terhadap variabel terikat.
2. Menentukan taraf nyata ( $\alpha=5\%$ ) dan derajat kebebasan ( $db$ ) untuk mengetahui nilai  $t_{tabel} = t_{(\alpha,db)}$ . Dimana  $n$  dan  $k$  masing-masing adalah jumlah contoh dan jumlah variabel bebas.
3. Mencari nilai  $t_{hitung}$  dinyatakan dengan rumus sebagai berikut (Ariefianto, 2012):

$$t_{hitung} = \frac{(\hat{\beta}_j)}{se(\hat{\beta}_j)} \quad (19)$$

dengan:

$\hat{\beta}_j$  : penduga parameter ke- $j$  yang dihipotesiskan  
 $se(\hat{\beta}_j)$  : Standard Error  $\hat{\beta}_j$

4. Kriteria Penolakan  
 $|t_{hitung}| \leq t_{(\alpha,db)}$ , berarti  $H_0$  diterima dan  $H_1$  ditolak, artinya tidak ada pengaruh antara variabel bebas terhadap variabel terikat.  
 $|t_{hitung}| > t_{(\alpha,db)}$ , berarti  $H_0$  ditolak dan  $H_1$  diterima, artinya ada pengaruh antara variabel bebas terhadap variabel terikat.

## 2.8. Koefisien Determinasi

Sebagai ukuran kebaikan model (*goodness of fit*) dapat menggunakan Jumlah Kuadrat Total (JKT) dan dekomposisinya serta koefisien  $R^2$ . JKT dan dekomposisinya dapat dihitung dengan jalan (Ariefianto, 2012):

$$\begin{aligned} \text{Jumlah kuadrat Total (JKT)} & \quad (20) \\ &= \sum_{i=1}^N \sum_{t=1}^T (y_{it} - \bar{y})^2 \end{aligned}$$

Jumlah Kuadrat yang dijelaskan (JKP) (21)

$$= \sum_{i=1}^N \sum_{t=1}^T (\hat{y}_{it} - \bar{y})^2$$

Jumlah Kuadrat Error (JKE) (22)

$$\begin{aligned} &= \sum_{i=1}^N \sum_{t=1}^T (y_{it} \\ &\quad - \hat{y}_{it})^2 \left( = \sum_{i=1}^N \sum_{t=1}^T u_{it}^2 \right) \end{aligned}$$

JKT adalah ukuran variasi contoh  $y_{it}$  untuk menunjukkan seberapa besar dispersi contoh  $y_{it}$  di sekitar rata-ratanya. JKP menunjukkan variasi contoh pada  $y$  dan JKE mengukur variasi dari  $u_{it}$ . Total variasi pada  $y$  adalah sama dengan jumlah JKP dan JKE, atau

$$JKT = JKP + JKE \quad (23)$$

selanjutnya, persamaan diatas dibagi dengan JKT, diperoleh

$$1 = \frac{JKP}{JKT} + \frac{JKE}{JKT} \quad (24)$$

sehingga koefisien determinasi,  $R^2$  dapat didefinisikan sebagai

$$\begin{aligned} R^2 &= \frac{JKP}{JKT} = 1 - \frac{JKE}{JKT} \\ &= 1 - \frac{\sum_{i=1}^N \sum_{t=1}^T u_{it}^2}{\sum_{i=1}^N \sum_{t=1}^T (y_{it} - \bar{y})^2} \end{aligned} \quad (25)$$

dari persamaan (25) dapat dilihat bahwa koefisien determinasi menunjukkan proporsi variasi variabel terikat ( $y$ ) yang dapat dijelaskan oleh variasi variabel bebas ( $x$ ). Nilai koefisien determinasi  $R^2$  mempunyai interval mulai dari 0 sampai 1 ( $0 \leq R^2 \leq 1$ ) karena nilai JKP dan JKE tidak mungkin melebihi nilai JKT.  $R^2$  adalah suatu ukuran kesesuaian model (*model fit*) (Sulaiman, 2004).

## 3. METODOLOGI PENELITIAN

### 3.1. Objek Penelitian

Objek penelitian dalam penelitian ini adalah data ketahanan pangan di setiap Kabupaten/Kota di Provinsi Bengkulu. Data tersebut diperoleh dari Dinas Pertanian Provinsi Bengkulu pada tahun 2009-2014.

<sup>1</sup>Alumni Jurusan Matematika FMIPA Universitas Bengkulu

<sup>2,3</sup>Staf Pengajar Jurusan Matematika FMIPA Universitas Bengkulu

### 3.2. Sumber Data

Data yang digunakan dalam penelitian adalah data sekunder, data tentang ketahanan pangan tahunan dari tahun 2009 hingga 2014, serta data silang sebanyak sepuluh kabupaten/kota di provinsi Bengkulu. Sumber data ialah berupa catatan dan laporan historis yang tersusun dalam arsip yang dipublikasikan dan tidak dipublikasikan yang diperoleh dari Dinas Pertanian Provinsi Bengkulu.

### 3.3. Variabel Penelitian

Variabel-variabel yang digunakan dalam penelitian ini adalah sebagai berikut:

- a. Rasio ketersediaan beras ( $Y$ )  
Rasio ketersediaan beras adalah angka perbandingan dari jumlah produksi beras dan konsumsi beras di tiap kabupaten/kota di Provinsi Bengkulu. Secara sistematis dapat ditulis
 
$$\text{Rasio ketersediaan beras} = \frac{\text{jumlah konsumsi beras}}{\text{jumlah produksi beras}}$$
- b. Ketersediaan beras ( $X_1$ )  
Ketersediaan beras merupakan jumlah beras yang dapat disimpan oleh suatu daerah disetiap tahun dalam satuan ton.
- c. Luas panen ( $X_2$ )  
Luas panen adalah jumlah areal yang dapat memproduksi padi setiap tahunnya dalam satuan hektar.
- d. Produktivitas lahan ( $X_3$ )  
Produktivitas lahan diukur berdasarkan rata-rata produksi padi yaitu rata-rata jumlah padi yang dapat dihasilkan dari 1 hektar lahan per tahun dalam satuan ton/hektar.
- e. Jumlah konsumsi beras ( $X_4$ )  
Jumlah konsumsi beras adalah jumlah beras yang dikonsumsi seluruh penduduk suatu kabupaten/kota dalam jangka waktu satu tahun dalam satuan ton.

- f. Harga beras ( $X_5$ )

Harga beras adalah harga komoditi beras yang sudah ditambah dengan biaya transportasi dalam pendistribusiannya (harga pasar) dalam satuan rupiah/kilogram.

### 3.4. Metode Analisis Data

Metode analisis yang akan digunakan adalah metode *standard error* Newey West untuk mengatasi heteroskedastisitas dan autokorelasi pada analisis regresi data panel, dengan tahapan meregresikan data adalah sebagai berikut:

1. Menentukan data panel (gabungan runtun waktu dan data silang ) yang akan digunakan dalam studi kasus.
2. Mengestimasi parameter regresi data panel dengan menggunakan *software Eviews*.
3. Menguji asumsi klasik untuk mendeteksi keberadaan heteroskedastisitas dan autokorelasi.
4. Melakukan transformasi
5. Menghitung nilai *standard error* Newey West
6. Membandingkan nilai *standard error* Newey West dengan nilai *standard error* OLS.
7. Menghitung kebaikan model dengan menggunakan *Eviews* 8 dan Ms.Excel.

## 4. HASIL DAN PEMBAHASAN

### 4.1. Deskripsi Data

Penelitian ini menggunakan data panel seimbang dengan jumlah sampel wilayah sebanyak 10 kabupaten/kota dari tahun 2009-2014. Data ini merupakan data sekunder yang diambil dari Dinas Pertanian, berikut variabel dalam penelitian.

Tabel 2. Variabel penelitian

simbol	Nama Variabel	Satuan
$Y$	Rasio Ketersediaan Beras	
$X_1$	Ketersediaan Beras	Ton
$X_2$	Luas Panen	Hektar
$X_3$	Harga Beras	Ton/ha
$X_4$	Jumlah Konsumsi Beras	Ton
$X_5$	Harga Beras	Rupiah

<sup>1</sup>Alumni Jurusan Matematika FMIPA Universitas Bengkulu

<sup>2,3</sup>Staf Pengajar Jurusan Matematika FMIPA Universitas Bengkulu

sedangkan wilayah yang digunakan dalam penelitian adalah kabupaten/kota yang ada di provinsi Bengkulu, seperti pada Tabel 3. berikut ini:

Tabel 3. Wilayah penelitian

No.	Wilayah
1	Bengkulu Selatan
2	Rejang Lebong
3	Bengkulu Utara
4	Kaur
5	Seluma
6	Muko-muko
7	Lebong
8	Kepahiang
9	Bengkulu Tengah
10	Kota Bengkulu

#### 4.2. Pemilihan Model Pendugaan

Pengolahan data menggunakan regresi data panel dengan alternatif tiga model, yaitu Model Efek Gabungan, Model Efek Tetap, dan Model Efek Acak. Dalam menentukan penduga model regresi data panel, hal pertama yang harus dilakukan adalah menguji model manakah yang paling tepat digunakan. Berikut pengujianya:

##### 4.2.1. Uji Chow

Statistik Uji

$$F = \frac{[R_{LSDV}^2 - R_{pooled}^2]/(N - 1)}{(1 - R_{LSDV}^2)/(NT - N - K)}$$

dengan jumlah data silang sebanyak 10, data runtun waktu 6 dan variabel bebas 5, maka jumlah data panel sebanyak 300, nilai dari  $R_{LSDV}^2$  dan  $R_{pooled}^2$  dapat dilihat pada lampiran output Model Efek Gabungan dan Model Efek Tetap maka secara manual nilai  $F$  dapat dihitung sebagai berikut:

$$F = \frac{[R_{LSDV}^2 - R_{pooled}^2]/(N - 1)}{(1 - R_{LSDV}^2)/(NT - N - K)}$$

$$F = \frac{\frac{0.919113 - 0.853494}{10 - 1}}{1 - 0.919113}$$

$$F = \frac{0.065619}{0.080887} = 4,056214$$

Kesimpulan

Nilai  $F_{hitung} = 4,056214 > F_{tabel} = 1,912924$ , maka  $H_0$  ditolak artinya model

yang tepat menggunakan Model Efek Tetap.

#### 4.2.2. Uji Hausman

Tabel 4. Tabel Uji husman

Test Summary	Chi-Sq.	
	Statistic	$p - value$
Cross-section random	4.817025	0.4386

Kesimpulan

Hasil dari uji Hausman dapat dilihat pada Tabel 4 dengan nilai  $p - value = 0,4386 > \alpha = 0,05$  maka  $H_0$  diterima, artinya model yang tepat adalah menggunakan Model Efek Tetap.

#### 4.3. Metode Pendugaan

Dari pemilihan model penduga diketahui bahwa model yang tepat digunakan adalah Model Efek Tetap. Namun, dalam penelitian ini diperlukan model yang memiliki masalah heteroskedastisitas dan autokorelasi ataupun salah satu dari keduanya. Untuk itu peneliti menggunakan model pertama yaitu Model Efek Gabungan. Model ini diketahui sebagai model yang tidak layak, sehingga diduga akan melanggar asumsi yang diberikan.

Tabel 5. Hasil Output Model Efek Gabungan

Variabel	Coefficient	Std. Error	t-statistic	$p - value$
$Y$	0.079396	0.442211	0.179542	0.8582
$X_1$	-1.21E-05	1.38E-05	-0.878808	0.3834
$X_2$	9.79E-05	2.94E-05	3.331809	0.0016
$X_3$	0.049208	0.010734	4.584270	0.0000
$X_4$	-5.78E-05	4.22E-06	-13.69486	0.0000
$X_5$	-2.90E-05	1.36E-05	-2.138462	0.0370
R-squared			0.853494	
F-statistic			62.91706	

Berdasarkan Tabel 5, diperoleh nilai penduga parameter, dan dapat ditulis dengan persamaan sebagai berikut:

$$y_{it} = 0.079396 + 0.0000121X_1 + 0.0000979X_2 + 0.049208X_3 - 0.0000578X_4 - 0.000029X_5 + \varepsilon_{it}$$

Taraf signifikansi yang digunakan dalam pengujian adalah 5%. Langkah

<sup>1</sup>Alumni Jurusan Matematika FMIPA Universitas Bengkulu

<sup>2,3</sup>Staf Pengajar Jurusan Matematika FMIPA Universitas Bengkulu

selanjutnya adalah melakukan pembentukan model regresi. Nilai penduga parameter dapat dilihat pada Tabel 5. dari tabel tersebut terlihat bahwa luas panen, produktivitas lahan, jumlah konsumsi beras dan harga beras berpengaruh terhadap rasio ketersediaan beras. Variabel ketersediaan beras tidak berpengaruh terhadap rasio ketersediaan beras karena  $p - value = 0,3834 > 0,05$ .

#### 4.4. Hasil Pengujian Asumsi Klasik

##### 4.4.1. Heteroskedastisitas

Dalam penelitian ini pengujian heteroskedastisitas yang digunakan adalah uji White. Berikut tabel uji White

Tabel 6. Uji Heteroskedastisitas White

	<i>P - value</i>	
F-statistic	2.562903F(5,54)	0.0375
Obs*R-squared	11.50754Chi-Square(5)	0.0422
Scaled explained SS	10.39682Chi-Square(5)	0.0647

Pada hasil pengujian diperoleh  $p - value obs * R - square = 0,0422 < 0,05$  sehingga dapat disimpulkan terdapat masalah heteroskedastisitas.

##### 4.4.2. Autokorelasi

Pada output regresi dengan Model Efek Gabungan diperoleh nilai  $DW - stat = 1,245781$ , dengan signifikansi 5%, nilai  $d$  bawah (*lower*) adalah 1,4083 dan  $d$  atas (*upper*) adalah 1,7671.

Tabel 7. Uji statistik Durbin-Watson (DW) dengan 5 variabel bebas

Nilai statistik $d$	Hasil
$0 \leq d \leq 1,4083$	$H_0$ ditolak: ada autokorelasi positif
$1,4083 \leq d \leq 1,7671$	Daerah keragu-raguan: tidak ada keputusan
$2 \leq d \leq 2,2329$	Gagal menolak $H_0$ tidak ada autokorelasi
$2,2329 \leq d \leq 2,5917$	Daerah keragu-raguan: tidak ada keputusan
$2,5917 \leq d \leq 4$	$H_0$ ditolak: ada autokorelasi negatif

terlihat nilai statistik Uji DW berada pada kisaran ( $0 \leq d \leq 1,4083$ ), sehingga dapat disimpulkan bahwa pada model tersebut mempunyai masalah autokorelasi..

#### 4.5. Standard Error Newey West

Selanjutnya dilakukan transformasi pada data yang bertujuan untuk mengatasi masalah yang terkandung didalamnya seperti masalah kenormalan. *Standard error* hasil regresi yang ditransformasi adalah tidak bias dan dengan demikian prosedur pengujian, baik uji  $t$  dan uji  $F$  menjadi valid. Berikut hasil transformasi tersebut:

Tabel 8. Penduga parameter dan pengujian data transformasi

Parameter	Nilai penduga	Std. Error	t-hitung	<i>p-value</i>
$\beta_0$	-2.667342	1.111548	-2.399664	0.0199
$\beta_1$	0.154808	0.205556	0.753118	0.4546
$\beta_2$	0.845985	0.200678	4.215630	0.0001
$\beta_3$	0.971097	0.232323	4.179945	0.0001
$\beta_4$	-0.943021	0.041899	-22.50717	0.0000
$\beta_5$	-0.095017	0.072404	-1.312314	0.1950

Variabel-variabel yang diuji adalah ketersediaan beras ( $X_1$ ), luas panen ( $X_2$ ), produktivitas lahan ( $X_3$ ), jumlah konsumsi beras ( $X_4$ ), dan harga beras ( $X_5$ ). Pada Tabel 8 dapat dilihat nilai  $p-value$  untuk masing-masing variabel. Berdasarkan pengujian uji  $t$  dengan membandingkan nilai  $p-value$  dan nilai  $\alpha = 5\%$  disimpulkan bahwa ketersediaan beras ( $X_1$ ), dan harga beras ( $X_5$ ) tidak berpengaruh terhadap rasio ketersediaan beras.

Tabel 9. Penduga parameter dan pengujian data dengan *Standard Error Newey West*  $g = 1$

Parameter	Nilai penduga	Std. Error	t-hitung	<i>p-value</i>
$\beta_0$	-2.667342	1.179538	-2.261344	0.0278
$\beta_1$	0.154808	0.091138	1.698612	0.0951
$\beta_2$	0.845985	0.083947	10.07765	0.0000
$\beta_3$	0.971097	0.098303	9.878644	0.0000
$\beta_4$	-0.943021	0.028855	-32.68179	0.0000
$\beta_5$	-0.095017	0.069395	-1.369210	0.1766

Pada Tabel 9 digunakan nilai  $g = 1$  karena data yang digunakan adalah data

<sup>1</sup>Alumni Jurusan Matematika FMIPA Universitas Bengkulu

<sup>2,3</sup>Staf Pengajar Jurusan Matematika FMIPA Universitas Bengkulu

tahunan sehingga dianjurkan untuk menggunakan nilai  $g = 1$ . Berdasarkan pengujian uji t dengan membandingkan nilai  $p$ -value dan nilai  $\alpha = 5\%$  disimpulkan bahwa ketersediaan beras ( $X_1$ ), dan harga beras ( $X_5$ ) tidak berpengaruh terhadap rasio ketersediaan beras. Sedangkan untuk Tabel 10 menggunakan nilai  $g = 4$ , nilai ini diperoleh dari rumus  $g = 4 \left(\frac{60}{100}\right)^{\frac{2}{9}} \approx 4$ .

Tabel 10. Penduga parameter dan pengujian data dengan *Standard Error* Newey West  $g = 4$

Para meter	Nilai penduga	Std. Error	t-hitung	$p$ -value
$\beta_0$	-2.667342	1.252164	-2.130185	0.0377
$\beta_1$	0.154808	0.069158	2.238479	0.0293
$\beta_2$	0.845985	0.061947	13.65652	0.0000
$\beta_3$	0.971097	0.095905	10.12561	0.0000
$\beta_4$	-0.943021	0.026447	-35.65657	0.0000
$\beta_5$	-0.095017	0.070185	-1.353802	0.1814

Setelah *standard error* dikoreksi dengan metode *standard error* Newey West hanya harga beras ( $X_5$ ) yang tidak berpengaruh terhadap rasio ketersediaan beras. Adanya heteroskedastisitas dan autokorelasi mengakibatkan statistik uji menjadi lebih besar. Oleh karena itu, keputusan yang dihasilkan dalam uji t akan lebih sering menyimpulkan adanya pengaruh variabel bebas terhadap variabel terikat.

Tabel 11. Ringkasan *Standard Error*

Parame ter	<i>Standard Error</i>		
	Data Transformasi	Newey West $g = 4$	Newey West $g = 1$
$\beta_0$	1.111548	1.252164	1.179538
$\beta_1$	0.205556	0.069158	0.091138
$\beta_2$	0.200678	0.061947	0.083947
$\beta_3$	0.232323	0.095905	0.098303
$\beta_4$	0.041899	0.026447	0.028855
$\beta_5$	0.072404	0.070185	0.069395

Berdasarkan Tabel 11 dapat dilihat bahwa *standard error* Newey West yang didapatkan, baik menggunakan  $g = 4$ ,

$g = 4$  selalu lebih kecil dari *standard error* data transformasi awal. Hal ini bahwa adanya autokorelasi akan menyebabkan *standard error* yang didapatkan dari metode OLS jauh dari nilai *standard error* yang sebenarnya.

Sebelum dilakukan koreksi terhadap *standard error* luas panen, produktivitas lahan, dan jumlah konsumsi beras berpengaruh terhadap rasio ketersediaan beras. Keberadaan autokorelasi dan heteroskedastisitas menyebabkan *standard error* lebih besar dari *standard error* sebenarnya. Namun setelah dilakukan koreksi terhadap *standard error* menggunakan metode *standard error* Newey West dengan dua pilihan nilai  $g$ , variabel yang berpengaruh ialah luas panen, produktivitas lahan, jumlah konsumsi beras dan harga beras berpengaruh terhadap rasio ketersediaan beras.

#### 4.6. Uji Signifikansi

Uji ini dilakukan untuk melihat apakah variabel-variabel bebas secara simultan maupun secara parsial berpengaruh terhadap variabel terikat.

##### 4.6.1. Uji Simultan (Uji F)

Nilai statistik uji F dapat dilihat pada Lampiran 8 nilai statistik uji F yang diperoleh sebesar 367,2679.

$$F(\alpha; k - 1; n - k) = F(5\%; 4; 55) = 2,539689$$

Kesimpulan

Dari pengujian diperoleh nilai  $|F| = 367,2679 > F_{tabel} = 2,539689$  sehingga  $H_1$  diterima, sehingga dapat dikatakan bahwa variabel bebas secara simultan berpengaruh terhadap variabel terikat.

##### 4.6.2. Uji Parsial (Uji t)

Perhatikan Tabel 12 dapat dilihat nilai  $t_{hitung}$  untuk masing-masing variabel. Berikut pengujian dari masing-masing variabel bebas dengan variabel

<sup>1</sup>Alumni Jurusan Matematika FMIPA Universitas Bengkulu

<sup>2,3</sup>Staf Pengajar Jurusan Matematika FMIPA Universitas Bengkulu

terikat dengan  $t_{tabel}: t(0,05; 55) = 2,004045$ .

Tabel 12. Uji Parsial

Parameter	$ t_{hitung} $	$t_{tabel}$	Keputusan
$\beta_1$	2.23848		$H_0$ ditolak
$\beta_2$	13.65652		$H_0$ ditolak
$\beta_3$	10.12561	2,004045	$H_0$ ditolak
$\beta_4$	35.65657		$H_0$ ditolak
$\beta_5$	1.35380		$H_0$ diterima

Kesimpulan:

Pada Tabel 12. yang membandingkan  $|t_{hitung}|$  dan  $t_{tabel}$  diperoleh keputusan bahwa pada pengujian hipotesis variabel ketersediaan beras, luas panen, produktivitas lahan, dan jumlah konsumsi beras  $H_0$  ditolak. Hal ini berarti masing-masing variabel secara parsial berpengaruh terhadap rasio ketersediaan beras. Sedangkan pada pengujian hipotesis variabel harga beras  $H_0$  diterima. Hal ini berarti variabel harga beras secara parsial tidak berpengaruh terhadap rasio ketersediaan beras.

#### 4.7. Pembentukan Model Terbaik

Berdasarkan pengujian yang telah dilakukan diketahui bahwa variabel  $X_5$  secara parsial tidak berpengaruh terhadap variabel terikat. Untuk itu dilakukan pengujian lebih lanjut dengan menghapus variabel  $X_5$ . Berikut hasil pengujiannya dengan bantuan *software Eviews 8* (secara lengkap dapat dilihat pada lampiran 9):

Tabel 13. Penduga parameter dan pengujian data tanpa variabel  $X_5$

Parameter	Nilai penduga	$t_{hitung}$	$p - value$
$\beta_0$	-3.623587	-5.208977	0.0000
$\beta_1$	0.104009	1.117016	0.2688
$\beta_2$	0.895943	10.60801	0.0000
$\beta_3$	1.007885	11.97222	0.0000
$\beta_4$	-0.943903	-26.73675	0.0000

dengan melihat Tabel 13 diatas dapat diketahui variabel mana saja yang berpengaruh signifikan terhadap variabel terikat. Oleh karena itu dilakukan

pengujian kembali tanpa mengikuti sertakan variabel  $X_1$ .

Tabel 14. Penduga parameter dan pengujian data tanpa variabel  $X_1$  dan  $X_5$

Param eter	Nilai penduga	$t_{hitung}$	$p - value$
$\beta_0$	-3.874892	-4.790866	0.0000
$\beta_2$	0.996374	60.74789	0.0000
$\beta_3$	1.103702	9.830114	0.0000
$\beta_4$	-0.942757	-26.49309	0.0000

Pada Tabel 14 dapat dilihat tidak ada lagi variabel yang memiliki nilai  $p - value$  lebih dari nilai  $\alpha = 5\%$  sehingga dapat disimpulkan semua variabel bebas yaitu luas panen, produktivitas lahan dan jumlah konsumsi beras secara individu mempengaruhi rasio ketersediaan beras. Berdasarkan Tabel 14. diperoleh nilai penduga parameter  $\beta_0 = -3.874892$ ,  $\beta_2 = 0.996374$ ,  $\beta_3 = 1.103702$  dan  $\beta_4 = -0.942757$ , dapat ditulis dengan persamaan sebagai berikut:

$$y_{it} = -3.874892 + 0.996374X_2 + 1.103702X_3 - 0.942757X_4 + \varepsilon_{it}$$

#### 4.8. Koefisien Determinasi

Perubahan pada *standard error* tidak mengubah nilai  $R^2$ , dari hasil penduga data transformasi didapatkan nilai  $R^2 adjusted$  sebesar 0.970382. Jadi keragaman rasio ketersediaan beras yang dapat dijelaskan oleh variabel bebas sebesar 97,038% sedangkan sisanya sebesar 2,962% dijelaskan oleh faktor lainnya.

## 5. KESIMPULAN DAN SARAN

### 5.1. Kesimpulan

Berdasarkan hasil pembahasan pada bab IV dapat disimpulkan bahwa analisis regresi data panel pada masalah ketahanan pangan ini menggunakan Model Efek Gabungan. Ini didasarkan pada pengujian telah dilakukan pemilihan model penduga dengan menggunakan uji Chow dan uji Hausman. Berdasarkan pengujian

<sup>1</sup>Alumni Jurusan Matematika FMIPA Universitas Bengkulu

<sup>2,3</sup>Staf Pengajar Jurusan Matematika FMIPA Universitas Bengkulu

diketahui bahwa Model Efek Gabungan merupakan model yang paling banyak melanggar asumsi. Hal ini terbukti dari pengujian asumsi diketahui bahwa data mengandung masalah heteroskedastisitas dan autokorelasi. Selanjutnya dilakukan pengujian dengan menggunakan metode *standard error* Newey West diperoleh variabel-variabel bebas yaitu ketersediaan beras, luas panen, produktivitas lahan, jumlah konsumsi beras yang mempengaruhi rasio ketersediaan beras sebagai variabel terikat.

## 5.2. Saran

Berdasarkan hasil penelitian, penulisan ingin menyarankan untuk penelitian selanjutnya, diharapkan dapat mencoba untuk menguji penelitian yang lebih dalam dengan menggunakan model lain yang ada pada regresi data panel.

## DAFTAR PUSTAKA

- Apriliawan, D., Tarno., Yasin, H. 2013. Pemodelan Laju Inflasi di Provinsi Jawa Tengah Menggunakan Regresi Data Panel. Hal. 301-321. *Jurnal Gaussian*. Vol.2, No.4. Universitas Diponegoro. Semarang.
- Ariefianto, M. D. 2012. *Ekonometrika: Esensi dan Aplikasi dengan Menggunakan Eviews*. Jakarta. Penerbit Erlangga.
- Azizah. 2013. Pengaruh Indeks Pembangunan Manusia, PDRB perkapita dan Tingkat Pengangguran Terhadap Jumlah Penduduk Miskin di Provinsi Jambi. *Jurnal Ilmiah Universitas Batanghari Jambi*. Vol.13, No.1. Universitas Batanghari Jambi.
- Dewi, H. R., dan D. E. Kusri. 2014. Peramalan Jumlah Kepemilikan Sepeda Motor dan Penjualan Sepeda Motor di Jawa Timur dengan Menggunakan Regresi Data Panel. *Jurnal Sains dan Seni Pomits*. Vol.03, No.02. Institut Teknologi Sepuluh November (ITS).
- Gujarati. 2004. *Basic Econometrics 4<sup>th</sup> edition*. New York. The McGraw-Hill Companies.
- Hsiao, C. 2014. *Analysis of Data Panel 3<sup>th</sup> edition*. New York. Cambridge University Press.
- Kutner, M. H. 2004. *Applied Linear Statistical Models 5<sup>th</sup> edition*. New York. McGraw-Hill Companies, Inc.
- Maharani, K., R. Fitriani., H. Pramoedyo. 2013. Kepekaan Metode *White Robust Standard Error* Dalam Mengatasi Masalah Heteroskedastisitas Pada Linier dan Kuadratik Pada Analisis Regresi Linier Berganda. *Jurnal Mahasiswa Statistik*. Vol.1, No.2. Universitas Brawijaya.
- Maruddani, D. A. I. 2014. *Model Praktikum Ekonometrika*. Semarang. Jurusan Statistika Fakultas Sains dan Matematika Universitas Diponegoro.
- Nurhayati, A. H. 2014. Pemodelan Balita Gizi Buruk di Kabupaten/Kota di Jawa Timur Tahun 2008 hingga 2011 dengan Analisis Regresi Panel. *Jurnal Mahasiswa Statistik*. Vol. 2, No.02. Universitas Brawijaya.
- Putri, N.U., Maiyastri., H. Yozza. 2013. Permasalahan Autokorelasi pada Analisis Regresi Linier Sederhana. *Jurnal Matematika*. Vol.2, No.2 : 26-34. Universitas Andalas.
- Rachmawati, D. S., E. Sumarminingsih. 2014. Metode Standard Error Newey West untuk mengatasi Heteroskedastisitas dan Autokorelasi pada Analisis Regresi Linier Berganda. *Jurnal Mahasiswa Statistik*. Vol.2, No.1 Universitas Brawijaya.
- Ratnasari, N. P. A. M., I. P. E. N. Kencana., G. K. Gandhiadi. 2014. Aplikasi Regresi Data Panel dengan

<sup>1</sup>Alumni Jurusan Matematika FMIPA Universitas Bengkulu

<sup>2,3</sup>Staf Pengajar Jurusan Matematika FMIPA Universitas Bengkulu

- Pendekatan FEM (Studi Kasus: PT. PLN Gianyar). *E-Jurnal Matematika*. Vol.3, No.1. Universitas Udayana.
- Silalahi, D., Sitepu, R., Tarigan G. 2014. Analisis Ketahanan Pangan Provinsi Sumatera Utara dengan Metode Regresi Data Panel. *Saintia Matematika*. Vol.02, No.03 : 237-251. Universitas Sumatera Utara. Medan.
- Sulaiman, W. 2004. *Analisis Regresi Menggunakan SPSS*. Yogyakarta. ANDI.
- Supardi, A. 2013. *Pertumbuhan Penduduk dan Ketahanan Pangan Provinsi Bengkulu*.  
<http://bengkulu.bkkbn.go.id/layout/s/mobile/dispform.aspx?List=8c526a76-8b88-44fe-8f81-2085df5b7dc7&View=69dc083c-a8aa-496a-9eb7-b54836a53e40&ID=79> (diakses pada 11 April 2015)
- Wooldridge, J. M. 2009. *Introductory Econometrics 4<sup>th</sup> Edition*. Canada. Nelson Education.

<sup>1</sup>Alumni Jurusan Matematika FMIPA Universitas Bengkulu

<sup>2,3</sup>Staf Pengajar Jurusan Matematika FMIPA Universitas Bengkulu

# MODEL KETAHANAN PANGAN PROVINSI BENGKULU DENGAN ANALISIS REGRESI DATA PANEL

Deqi Suanto<sup>1)</sup>, Sigit Nugroho<sup>2)</sup>, Idhia Sriliana<sup>2)</sup>

<sup>1)</sup> Alumni Jurusan Matematika FMIPA Universitas Bengkulu

<sup>2)</sup> Staf Pengajar Jurusan Matematika FMIPA Universitas Bengkulu

---

## ABSTRAK

Penelitian ini bertujuan untuk mengetahui model ketahanan pangan di Provinsi Bengkulu dengan menggunakan analisis regresi data panel. Data yang digunakan dalam penelitian ini yaitu data ketahanan pangan beras dari 10 wilayah kabupaten/kota di Provinsi Bengkulu dari tahun 2009-2015. Data panel adalah data yang terdiri dari beberapa periode waktu dan beberapa objek. Berdasarkan penelitian diperoleh model terbaik untuk ketahanan pangan Provinsi Bengkulu adalah *fixed effect model*. Variabel yang signifikan terhadap ketahanan pangan yaitu luas areal panen padi ( $x_2$ ) dan produktivitas lahan ( $x_3$ ). Sedangkan, variabel yang tidak signifikan yaitu stok beras ( $x_1$ ), jumlah konsumsi beras ( $x_4$ ) dan harga beras ( $x_5$ ).

**Kata Kunci :** Ketahanan Pangan, *Time Series*, *Cross-Section*, Regresi Data Panel

---

## I. PENDAHULUAN

Sebagai salah satu produsen bahan pangan dunia, krisis pangan internasional turut berpengaruh pada persediaan pangan di Indonesia. Menurut Badan Pusat Statistik (BPS), pada Agustus 2015 nilai ekspor Indonesia mencapai US\$12,70 miliar atau meningkat 10,79 persen dibanding ekspor Juli 2015. Namun, tingginya permintaan akan bahan pangan tidak seimbang dengan persediaan alam Indonesia dan mengakibatkan kelangkaan pangan dalam negeri. Persediaan beras sebagai bahan pangan pokok juga memperlihatkan penurunan signifikan. Menurut seorang guru besar Fakultas Pertanian IPB, Andreas Santosa bahwa stok beras Januari 2013 mencapai 7,4 juta ton, mengalami penurunan pada 2014 menjadi 6,5 juta ton, menyisakan 5,5 juta ton pada Januari 2015, dan menurun 15% pada akhir tahun 2015. Kelangkaan pangan juga menyebabkan Indonesia harus mengimpor beras dari negara penghasil beras lain seperti Thailand dan India (Anonim, 2015).

Menurut Badan Pusat Statistik (BPS) Provinsi Bengkulu, produksi padi Provinsi Bengkulu tahun 2015 sebanyak 578.654 ton Gabah Kering Giling (GKG), dibandingkan produksi padi tahun 2014 yang mencapai 593.194 ton mengalami penurunan sebanyak 14.540 ton atau -2,45 persen. Produksi padi tahun 2014 juga mengalami penurunan sebesar 4,76 persen (29.637ton) dibanding tahun 2013 yang mencapai 622.832 ton. Provinsi Bengkulu

mengalami penurunan produksi padi setiap tahunnya sehingga ketahanan pangan beras yang bersumber dari produksi padi menjadi berkurang disamping pertumbuhan penduduk yang semakin bertambah semakin meningkat pula kebutuhan pangan. Oleh sebab itu, pemerintah harus melaksanakan kebijakan pangan yaitu menjamin ketahanan pangan Provinsi Bengkulu dengan meningkatkan produksi padi disetiap tahunnya agar kebutuhan pangan dan pertumbuhan penduduk menjadi seimbang.

Pada penelitian ini penulis menggunakan metode regresi data panel untuk menganalisis ketahanan pangan di Provinsi Bengkulu. Metode ini dipilih karena data yang digunakan terdiri dari data *cross-section* yaitu data ketersediaan beras dari 10 kabupaten yang ada di Provinsi Bengkulu dan data *time series* berupa periode waktu yang digunakan dari tahun 2009-2015. Sehingga judul untuk penelitian ini adalah "*Model Ketahanan Pangan Provinsi Bengkulu Dengan Analisis Regresi Data Panel*".

## II. LANDASAN TEORI

### A. Pengertian Ketahanan Pangan

FAO (*Food and Agriculture Organization*) mendefinisikan ketahanan pangan sebagai situasi dimana dalam segala waktu memiliki kecukupan jumlah atas pangan yang aman dan bergizi demi kehidupan yang sehat dan aktif. Secara umum ketahanan pangan adalah adanya jaminan bahwa kebutuhan

pangan dan gizi setiap penduduk adalah sebagai syarat utama dalam mencapai derajat kesehatan dan kesejahteraan yang tercukupi.

## **B. Faktor-faktor yang Mempengaruhi Ketahanan Pangan**

Faktor-faktor yang mempengaruhi ketahanan pangan adalah sebagai berikut (Silalahi, 2014) :

### 1. Stok Beras

Jumlah beras yang dapat disimpan setiap tahun dapat menjadi salah satu indikator ketahanan pangan. Semakin banyak beras yang dapat disimpan oleh suatu daerah, maka ketahanan pangan di daerah tersebut semakin baik. Menurut Bulog, tersedianya kebutuhan beras minimal untuk 3 bulan ke depan disuatu daerah, menjadi indikasi bahwa daerah tersebut dikatakan tahan pangan.

### 2. Luas Areal Panen Padi

Luas lahan pertanian akan mempengaruhi skala usaha, dan pada akhirnya skala usaha ini akan mempengaruhi efisien atau tidaknya suatu usaha pertanian. Sering dijumpai makin luas areal panen yang dipakai untuk pertanian akan semakin tidak efisien lahan tersebut. Sebaliknya luasan areal panen yang sempit, upaya pengusaha terhadap penggunaan faktor produksi semakin baik, penggunaan tenaga kerja yang tercukupi dan tersedianya modal yang tidak terlalu besar sehingga usaha pertanian yang seperti ini sering lebih efisien. Meskipun demikian luas areal panen yang terlalu kecil cenderung menghasilkan usaha yang tidak efisien.

### 3. Produktivitas Lahan Padi

Produktivitas adalah kemampuan suatu tanah untuk menghasilkan suatu tanaman yang sedang diusahakan dengan system pengelolaan tertentu. Keahlian tentang pertanian menjadi faktor yang sangat mempengaruhi produktivitas suatu lahan. Dapat dikatakan semakin berpendidikan petani-petani di suatu wilayah maka keberhasilan produksi akan semakin meningkat.

### 4. Jumlah Konsumsi Beras per Kapita

Pemerintah pusat maupun pemerintah daerah sedang menerapkan diversifikasi pangan yang diharapkan dapat mengurangi jumlah konsumsi beras. Ketergantungan yang sangat besar terhadap beras telah menggusur budaya makan pangan lokal yang beragam dan sudah teruji sejarah dan berlangsung berabad-abad.

Saat ini 95 persen perut penduduk Indonesia sangat tergantung pada makanan yang bernama nasi, sumbangan beras terhadap energi dan protein masih sangat tinggi, yaitu lebih dari 55%.

### 5. Harga Beras

Dalam upaya meningkatkan produktivitas, pemerintah membuat kebijakan terhadap harga beras yaitu Harga Pembelian Pemerintah (HPP). Kebijakan tersebut bertujuan agar petani padi menerima harga gabah yang layak, sehingga mereka menerima insentif untuk meningkatkan produktivitas. Penetapan HPP berdasarkan pertimbangan agar petani dapat menerima keuntungan minimal 28% dari harga yang diterima.

### 6. Rasio Ketersediaan Beras

Rasio ketersediaan beras merupakan angka perbandingan dari jumlah produksi beras dan jumlah konsumsi beras.

## **C. Jenis-jenis Data**

Data yang dilihat dari proses pengambilannya dapat dikelompokkan menjadi tiga yaitu data *time series*, *cross-section* dan data panel dengan penjelasan sebagai berikut :

1. Data *time series* biasanya meliputi suatu objek (misalnya tingkat inflasi, laba, investasi, pertumbuhan ekonomi, dll), tetapi meliputi beberapa periode (bisa harian, bulanan, kuartalan, tahunan dan sebagainya) yang karakteristik runtun waktunya dapat berubah-ubah seiring berjalannya waktu.
2. Data *cross-section* adalah data terdiri atas beberapa atau banyak objek, sering disebut responden (misalnya perusahaan, provinsi, kabupaten, negara, dll) dengan waktunya dalam satu periode waktu.
3. Data panel (*pooled data*) adalah jenis data gabungan antara data *time series* dan *cross-section*. Oleh karena itu data panel mempunyai penggabungan karakteristik dari *time series* dan data *cross-section* yaitu data yang terdiri dari beberapa periode waktu dan beberapa objek (Widarjono, 2007).

## **D. Regresi Data Panel**

Menurut Hsiao (2003) data panel merupakan gabungan data *time series* dan data *cross-section* dengan kata lain data panel merupakan data dari beberapa objek dan beberapa periode waktu tertentu.

Model regresi data panel secara umum dapat dinyatakan dalam bentuk berikut (Hsiao, 2003) :

$$Y_{it} = \alpha^* + \sum_{k=1}^K \beta_k x_{kit} + u_{it} \quad (2.1)$$

$i = 1, 2, \dots, N ; t = 1, 2, \dots, T, k = 1, 2, \dots, K$

dimana:

$Y_{it}$  = variabel terikat unit *cross-section* ke- $i$  untuk *time series* ke- $t$

$\beta_k = (\beta_1, \beta_2, \dots, \beta_k)$  adalah parameter koefisien regresi

$x_{kit} = (x_{1it}, x_{2it}, \dots, x_{kit})$  adalah menunjukkan observasi pada variabel bebas berukuran  $1 \times K$ , di mana  $K$  adalah banyaknya variabel bebas.

$\alpha^*$  = merupakan efek individu dari unit *cross-section* ke- $i$  dan waktu ke-

$u_{it}$  = *error* regresi untuk group ke- $i$  untuk periode ke- $t$ .

### E. Model Estimasi Parameter

Menurut Widarjono (2007), bahwa dalam mengestimasi model regresi data panel terdapat tiga pendekatan yang biasa digunakan yaitu sebagai berikut:

#### 1. Common Effect Model (CEM)

Pendekatan pertama ini merupakan paling sederhana yang disebut estimasi *common effect model* atau *pooled least square*. Pendekatan ini diasumsikan bahwa nilai intersep masing-masing variabel sama, begitu pula *slope* koefisien untuk semua unit *time series* dan *cross-section*. Berdasarkan asumsi ini maka persamaan dari *common effect model* adalah sebagai berikut:

$$Y_{it} = \alpha^* + \sum_{k=1}^K \beta_k x_{kit} + u_{it} \quad (2.2)$$

Persamaan diatas sama seperti dengan model dasar regresi data panel, dimana  $t$  merupakan periode waktu ( $t = 1, 2, \dots, T$ ),  $i$  merupakan jumlah individu ( $i = 1, 2, \dots, N$ ). Sehingga dengan metode *common effect model* sulit melihat perubahan antar individu karena semua dianggap sama (Gujarati, 2004).

#### 2. Fixed Effect Model (FEM)

Model ini mengasumsikan bahwa nilai intersep berbeda-beda di setiap unit tapi memiliki nilai yang sama di *slope*. Untuk mengestimasi data panel *fixed effect model* menggunakan teknik variabel dummy untuk menangkap perbedaan intersep antar *cross-section*. Namun demikian *slope*-nya sama antar objek. Model estimasi ini sering juga disebut dengan teknik *Least Square Dummy Variabel*

(LSDV). Model dari FEM dapat ditulis sebagai berikut:

$$Y_{it} = \alpha_i^* + \sum_{k=1}^K \beta_k x_{kit} + u_{it} \quad (2.3)$$

Perbedaan antar persamaan (2.2) dan (2.3) terletak pada intersep. Pada persamaan (2.3) indeks  $i$  menunjukkan bahwa intersep dari masing-masing *cross-section* berbeda. Menurut Gujarati (2004), perbedaan yang terjadi dikarenakan perbedaan karakteristik wilayah observasi berbeda dari satu ke yang lain.

#### 3. Random Effect Model (REM)

Model ini akan mengestimasi data panel dimana variabel mungkin saling berhubungan antar waktu dan antar individu. Keuntungan menggunakan *random effect model* yakni menghilangkan heterokedastisitas. Model ini juga disebut *Error Component Model* (ECM) atau teknik *Generalized Least Square* (GLS).

Menurut Gujarati (2004) pada model *random effect model* diasumsikan  $\alpha_i$  merupakan variabel *random* dengan  $mean \alpha_0$ , sehingga intersep dapat dinyatakan sebagai  $\alpha_i^* = \alpha + \varepsilon_i$  dengan  $\varepsilon_i$  merupakan *error random* yang mempunyai mean 0 dan varians  $\sigma_\varepsilon^2$ ,  $\varepsilon_i$  tidak secara langsung diobservasi atau disebut juga variabel laten. Jadi persamaan adalah sebagai berikut (Gujarati, 2004) :

$$Y = \alpha_i^* + \sum_{k=1}^K \beta_k x_{kit} + u_{it} \quad (2.4)$$

$$= \alpha^* + \sum_{k=1}^K \beta_k x_{kit} + \omega_{it}$$

Dengan  $\omega_{it} = \varepsilon_i + u_{it}$ . Suku *error* gabungan  $\omega_{it}$  memuat dua komponen *error* yaitu  $\varepsilon_i$  komponen *error cross-section* dan  $u_{it}$  yang merupakan komponen *error time series* dan *cross-section*. Karena inilah *random effect model* sering disebut juga *Error Components Model* (ECM). Asumsi yang berlaku pada *random effect model* adalah

- $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$ ,
- $u_{it} \sim N(0, \sigma_u^2)$
- $E(\varepsilon_i u_{it}) = 0, E(\varepsilon_i \varepsilon_j) = 0 (i \neq j)$ ,
- $E(u_{it} u_{is}) = E(u_{it} u_{jt}) = E(u_{it} u_{js}) (i \neq j; t \neq s)$ .

### F. Pemilihan Model Estimasi

Dalam menentukan estimasi model yang paling cocok pada regresi data panel, dilakukan beberapa uji untuk mendapatkan model yang paling sesuai.

#### 1. Uji Chow

Uji *Chow* digunakan untuk memilih kedua metode diantara *common effect model* dan *fixed effect model*.

## 2. Uji Hausman

Uji hausman digunakan untuk membandingkan *fixed effect model* dengan *random effect model*.

## G. Uji Asumsi Klasik

Suatu model dikatakan baik apabila model tersebut memenuhi beberapa asumsi yang disebut asumsi klasik.

### 1. Asumsi Normalitas

Uji normalitas digunakan untuk mengetahui apakah populasi data berdistribusi normal atau tidak.

### 2. Uji Multikolinieritas

Uji multikolinieritas adalah asumsi untuk mengetahui ada atau tidaknya korelasi yang signifikan antara variabel-variabel bebas dalam suatu model regresi linier berganda.

### 3. Asumsi Heteroskedastisitas

Uji heteroskedastisitas bertujuan untuk menguji apakah dalam model regresi terjadi ketidaksamaan varians residual, jika dari satu pengamatan ke pengamatan lain tetap, maka disebut homoskedastisitas dan jika berbeda disebut heteroskedastisitas.

### 4. Asumsi Autokorelasi

Uji autokorelasi digunakan untuk melihat apakah ada hubungan linier antara *error* serangkaian observasi yang di urutkan menurut waktu (data *time series*).

## H. Uji Signifikansi

Menurut Gujarati dan Porter (2010), uji signifikansi merupakan prosedur yang digunakan untuk menguji kebenaran atau kesalahan dari hasil hipotesis nol dari sampel.

### 1. Uji Individu

Menurut Statsdata (2011), uji individu (uji t) pada dasarnya menunjukkan seberapa jauh pengaruh variabel penjelas secara individual dalam menerangkan variabel-variabel terikat.

### 2. Uji Serentak

Uji serentak (Uji F) digunakan untuk menguji variabel-variabel bebas secara bersama terhadap variabel terikat.

### 3. Koefisien Determinasi

Ketepatan model  $R^2$  dilakukan untuk mendeteksi ketepatan yang paling baik dari garis regresi.

## III. METODE PENELITIAN

### A. Jenis Penelitian

Penelitian ini merupakan penelitian statistik terapan, yaitu suatu penelitian yang dilakukan dengan mengaplikasikan metode

statistika ke dalam bidang-bidang tertentu. Pada penelitian ini, metode regresi data panel digunakan untuk menganalisis pengaruh stok beras, luas panen padi, produktivitas lahan, jumlah konsumsi beras dan harga beras terhadap ketahanan pangan di Provinsi Bengkulu.

### B. Pengambilan Data

Data yang digunakan dalam penelitian ini merupakan data sekunder dari Dinas Pertanian Provinsi Bengkulu dan Badan Ketahanan Pangan Provinsi Bengkulu yaitu data rasio ketersediaan beras, stok beras, luas panen padi, produktivitas lahan, jumlah konsumsi beras dan harga beras setiap kabupaten/kota Provinsi Bengkulu dari tahun 2009-2015.

### C. Variabel Penelitian

Adapun variabel yang di gunakan dalam penelitian ini yaitu

1. Variabel Terikat : rasio ketersediaan beras ( $Y$ )
2. Variabel Bebas  
Stok beras ( $X_1$ ), luas panen padi ( $X_2$ ), produktivitas lahan ( $X_3$ ), jumlah konsumsi beras ( $X_4$ ) dan harga beras ( $X_5$ ).

### C. Tehnik Analisis Data

Metode analisis yang akan digunakan adalah regresi data panel. Tahapan yang akan digunakan dalam analisis data adalah sebagai berikut:

1. Pengumpulan data rasio ketersediaan beras, stok beras, luas panen padi, produktivitas lahan, jumlah konsumsi beras dan harga beras untuk setiap Kabupaten Kota Provinsi Bengkulu.
2. Mengidentifikasi variabel, dimana rasio ketersediaan beras sebagai variabel terikat ( $Y$ ) dan variabel bebasnya adalah stok beras ( $X_1$ ), luas panen padi ( $X_2$ ), produktivitas lahan ( $X_3$ ), jumlah konsumsi beras ( $X_4$ ) dan harga beras ( $X_5$ ).
3. Mendeskripsikan data penelitian dalam bentuk diagram batang.
4. Melakukan estimasi model regresi data panel dengan menggunakan program *views 8.0*.
5. Setelah itu, dengan menggunakan uji chow dan uji hausman akan dipilih model regresi terbaik menggunakan program *views 8.0*.
6. Setelah menentukan teknik estimasi, dilakukan uji asumsi klasik dan uji

signifikansi untuk menentukan model yang terbaik.

#### IV. HASIL DAN PEMBAHASAN

##### A. Deskripsi Data

Penelitian ini menggunakan data sekunder yaitu stok beras, luas areal panen padi, produktivitas lahan padi, jumlah konsumsi beras, harga beras dan rasio ketersediaan beras dari Tahun 2009 sampai Tahun 2015.

##### B. Estimasi Parameter

Estimasi parameter dilakukan dengan menggunakan *Common Effect Model*, *Fixed Effect Model* dan *Random Effect Model*.

###### 1. *CommonEffect Model*

*Common effect model* merupakan metode yang mengkombinasikan atau mengumpulkan semua data *cross section* dan *time series*.

Model ketahanan pangan dengan metode *common effect model* adalah sebagai berikut:

$$\hat{Y} = 0,431683 + 0,00000264 X_{1it} + 0,0000646X_{2it} + 0,000381X_{3it} - 0,0000540X_{4it} - 0,0000242X_{5it}$$

###### 2. *FixedEffect Model*

Model ini mengasumsikan bahwa nilai intersep berbeda-beda disetiap unit tapi memiliki nilai yang sama di slope. Untuk mengestimasi data panel *fixed effect model* menggunakan teknik variabel dummy untuk menangkap perbedaan intersep antar *cross-section* sehingga ada pendekatan dengan memperhatikan *cross-section*.

Model ketahanan pangan dengan metode *fixed effect model* sebagai berikut :

$$\hat{Y} = \alpha_i^* - 0,00000130 X_{1it} + 0,0000915X_{2it} + 0,000353X_{3it} + 0,0000123X_{4it} - 0,0000254X_{5it}$$

dimana :

$$\alpha_i^* = -1,025582 + c_i$$

$c_i$  = nilai intersep yang tidak diketahui bergantung pada wilayah

###### 3. *Random Effect Model*

Model ini akan mengestimasi data panel dimana variabel mungkin saling berhubungan antar waktu dan antar individu.

model ketahanan pangan dengan metode *random effect model* adalah sebagai berikut:

$$\hat{Y} = 0,178020 + 0,000000483X_{1it} + 0,0000724X_{2it} + 0,000386X_{3it} - 0,0000504X_{4it} - 0,0000114X_{5it}$$

#### V. PENUTUP

##### A. Kesimpulan

Berdasarkan hasil dan pembahasan, analisis regresi data panel dapat digunakan

dalam melihat pengaruh stok beras, luas panen padi, produktivitas lahan, jumlah konsumsi beras dan harga beras terhadap ketahanan pangan Provinsi Bengkulu. Kesimpulan dari penelitian ini adalah :

1. Model ketahanan pangan terbaik untuk Provinsi Bengkulu berdasarkan analisis regresi data panel yaitu *fixed effect model*.
2. Berdasarkan model terbaik terdapat dua faktor yang signifikan yaitu faktor luas areal panen padi dan produktivitas lahan.
3. Model ketahanan pangan di Provinsi Bengkulu dengan *fixed effect model* adalah sebagai berikut:

$$\hat{Y} = \alpha_i^* + 0,0000837X_{2it} + 0,000340X_{3it}$$

##### B. Saran

Berdasarkan hasil penelitian, penulis ingin memberikan saran atau masukan untuk penelitian selanjutnya, yaitu :

1. Diharapkan, mencoba untuk menganalisis yang lebih dalam dengan penambahan jumlah variabel bebas, *cross section* dan *time series*. Serta mampu menggunakan *software* yang berbeda dalam menghasilkan metode estimasi.
2. Penelitian ini dapat menggunakan metode yang berbeda untuk dianalisis bersama dengan regresi data panel, sehingga dapat menghasilkan metode terbaik.

#### DAFTAR PUSTAKA

- Aisyah, R. 2014. *Model Regresi Data Panel Persentase Penduduk Miskin Di Provinsi Bengkulu*. U nitas Bengkulu.
- Anonim, 2015. *Berita Resmi Statistik BPS Bengkulu*. URL: [http://bengkulu.bps.go.id/webbeta/websi te/brs\\_ind/brsInd20150330164709.pdf](http://bengkulu.bps.go.id/webbeta/websi te/brs_ind/brsInd20150330164709.pdf) (tanggal akses: 3 April 2016)
- Arifin, B. 2004. *Analisis Ekonomi Pertanian Indonesia*. Jakarta: Penerbit Buku Kompas.
- Baltagi, B. 2005. *Econometric Analysis of Panel Data*. Wiley
- Greene, W.H. 2007. *Econometric Analysis*. Sixth Edition. Prentice Hall. New York University.

- Gujarati, D. 2004. *Basic Econometrics*. Fourth Edition. McGraw-Hill. New York University.
- Gujarati, D. 2007. *Dasar-Dasar Ekonometrika Deret Waktu* Jilid. Jakarta: Penerbit Buku Erlangga.
- Gujarati, D.N. dan Porter, D.C. 2010. *Dasar-dasar Ekonometrika*. Jilid 1. Edisi Kelima. Mardanugraha, dkk [Penerjemah]. Jakarta: Salemba Empat.
- Hikam, M.A.S. 2014. *Memperkuat Ketahanan Pangan Demi Masa Depan Indonesia 2015-2025*. Jakarta Pusat. Penerbit Buku : cv. Rumah buku
- Hsiao, C. 2003. *Analysis of Data Panel*. Second Edition. Cambridge University Press
- Juanda, B dan Junaidi. 2012. *Ekonometrika Deret Waktu Teori dan Aplikasi*. Bogor: Penerbit IPB Press.
- Marsigit, W. 2010. *Pengembangan Diversifikasi Produk Pangan Olahan Lokal Bengkulu Untuk Menunjang Ketahanan Pangan Berkelanjutan*. Jurnal Teknologi Pertanian, Universitas Bengkulu. Bengkulu **Vol. 30, No. 4, November 2010**
- Puradisastra, M.D.N. 2006. *Analisis Ketahanan Pangan Kabupaten Nganjuk Berdasarkan Angka Kecukupan Energi Dan Pola Pangan Harapan Wilayah*. [Skripsi] Fakultas Pertanian IPB. Bogor.
- Silalahi, D. 2014. "Analisis Ketahanan Pangan Provinsi Sumatra Utara Dengan Metode Regresi Data Panel". *Jurnal Sainia Matematika*. **Vol. 02, No. 03, 21 April 2014**
- Statsdata. 2011. *Uji Asumsi Klasik Regresi Linier*. url: <http://www.statsdata.my.id/2011/12/uji-asmusi-klasik-regresi-linier.html> (tanggal akses: 3 April 2016)
- Sugandi, D. 2011. "Analisis Kebijakan Pembangunan Ketahanan Pangan Di Provinsi Bengkulu". *Laporan Akhir Badan Penelitian Dan Pengembangan Pertanian Kementerian Pertanian*. **Kode Registrasi 1801.15.018**
- Supranto, J. 2009. *Statistik Teori dan Aplikasi* Edisi Ketujuh. Erlangga. Jakarta.
- Widarjono, W. 2007. *Analisis ekonometrika dan Statistika dengan Eviws*. UPP STIM YKPN. Yogyakarta.
- Widowati, S. 2003. "Prospek Tepung Sukun Untuk Berbagai Produk Makanan Olahan dalam Upaya Menunjang Diversifikasi Pangan". *Makalah Pribadi. Pengantar Kefalsafah Sains Program Pasca Sarjana IPB Bogor*.

**MODEL LOGIT DAN MODEL PROBIT FAKTOR-FAKTOR KEBERHASILAN  
BELAJAR MAHASISWA PADA SISTEM PEMBELAJARAN JARAK JAUH (SPJJ)**  
(Studi Kasus: Mahasiswa Jurusan Manajemen Fakultas Ekonomi Universitas Terbuka (JME-  
UT))

**Mesi Yuliza<sup>1</sup> \*, Sigit Nugroho<sup>2</sup>, Dyah Setyo Rini<sup>3</sup>**  
Jurusan Matematika, FMIPA, Universitas Bengkulu  
Jl. WR. Supratman Kandang Limun Bengkulu 38371A  
e-mail : [mesiyuliza@gmail.com](mailto:mesiyuliza@gmail.com)<sup>1\*</sup>

**ABSTRAK**

Penelitian ini bertujuan untuk mengetahui faktor-faktor keberhasilan belajar mahasiswa pada sistem pembelajaran jarak jauh (SPJJ) mahasiswa JME-UT Bengkulu. Data yang digunakan dalam penelitian ini adalah data primer dan data sekunder. Data sekunder berupa data mahasiswa JME-UT yang sudah memiliki Indeks Prestasi Kumulatif (IPK) dan data primer berupa data yang dikumpulkan dari kuesioner yang dibagikan kepada 172 mahasiswa JME-UT. Model yang digunakan dalam penelitian ini adalah model logit dan model probit. Hasil penelitian ini diperoleh bahwa pada model logit terdapat dua faktor yang berpengaruh, yaitu motivasi dalam belajar dan jumlah sks, sedangkan pada model probit terdapat tiga faktor yang berpengaruh, yaitu umur, motivasi dalam belajar dan jumlah sks.

**Kata kunci:** *Keberhasilan belajar, SPJJ, Model Logit, Model Probit.*

**LOGIT MODEL AND PROBIT MODEL FACTORS OF STUDENT LEARNING IN  
THE DISTANCE LEARNING SYSTEM (SPJJ)**

(Case Study: Student Management Department Faculty of Economics of Open University  
(JME-UT))

**Mesi Yuliza<sup>1</sup> \*, Sigit Nugroho<sup>2</sup>, Dyah Setyo Rini<sup>3</sup>**  
Jurusan Matematika, FMIPA, Universitas Bengkulu  
Jl. WR. Supratman Kandang Limun Bengkulu 38371A  
e-mail : [mesiyuliza@gmail.com](mailto:mesiyuliza@gmail.com)<sup>1\*</sup>

**ABSTRACT**

This study aims to determine the factors of student success in distance learning system (SPJJ) JME-UT Bengkulu students. The data used in this research are primary data and secondary data. Secondary data are data of JME-UT students who already have Cumulative Achievement Index (GPA) and primary data in the form of data collected from questionnaires distributed to 172 JME-UT students. The model used in this research is logit model and probit model. The results of this study showed that there are two factors that influence the logit model, namely the motivation in learning and the number of credits, while in the probit model there are three influential factors, namely age, motivation in learning and the number of credits.

**Keywords:** *Learning success, SPJJ, Logit Model, Probit Model.*

## **A. PENDAHULUAN**

### **1. Latar Belakang**

Universitas Terbuka menerapkan sistem pembelajaran jarak jauh (SPJJ). Istilah jarak jauh berarti pembelajaran tidak dilakukan secara tatap muka, melainkan menggunakan media, baik media cetak (modul) maupun non-cetak (audio/video, komputer/internet, siaran radio, dan televisi). Pada sistem pembelajaran jarak jauh (SPJJ), mahasiswa diharapkan dapat belajar secara mandiri dan mahasiswa juga dituntut untuk mengatur waktunya secara teratur dan efisien. Tujuannya agar mahasiswa dapat mencapai hasil belajar yang optimal. Universitas Terbuka seperti halnya perguruan tinggi yang lain, menerapkan sistem kredit semester (sks) untuk menetapkan beban studi mahasiswa [2].

Karakteristik SPJJ berbeda dengan pendidikan informal, belajar sendiri, atau otodidak karena SPJJ mengharuskan adanya lembaga atau organisasi yang mengelola segala proses pendidikan. Oleh karena itu bila ada orang yang belajar sendiri dengan cara mengatur proses dan waktu sendiri tanpa adanya lembaga pendidikan yang mempengaruhinya, maka tidak dapat dikatakan sebagai SPJJ melainkan hanya sebagai otodidak atau studi pribadi [4]. Salah satu indikator keberhasilan belajar mahasiswa adalah Indeks Prestasi Kumulatif (IPK). Misalnya kategori "Berhasil" apabila  $IPK \geq 3$  dan kategori "Tidak Berhasil" apabila  $IPK < 3$  [8]. Ada beberapa metode yang digunakan dalam menentukan faktor-faktor yang mempengaruhi keberhasilan belajar mahasiswa pada SPJJ. Diantaranya model probit dan regresi logistik biner.

Regresi Logistik Biner merupakan suatu metode regresi logistik yang digunakan untuk mencari hubungan antara variabel respon ( $y$ ) yang bersifat biner atau dikotomis dengan variabel prediktor ( $x$ ) yang bersifat polikotomis [3]. Sedangkan model probit merupakan model nonlinier yang menggunakan bilangan biner (variabel boneka) sebagai variabel responnya. Variabel boneka yang dimaksud adalah jenis variabel diskrit yang mempunyai dua nilai yaitu 0 dan 1 [10]. Karena hanya memiliki dua nilai (0 dan 1) maka variabel respon identik dengan distribusi Bernoulli.

### **2. Rumusan Masalah**

Berdasarkan latar belakang di atas, adapun rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana menguji variabel-variabel yang diduga berpengaruh terhadap keberhasilan belajar mahasiswa pada SPJJ ?
2. Bagaimana model keberhasilan belajar mahasiswa pada SPJJ dengan menggunakan model logit dan model probit ?

### **3. Tujuan**

Tujuan yang ingin dicapai oleh penelitian ini adalah:

1. Menguji variabel-variabel yang diduga berpengaruh terhadap keberhasilan belajar mahasiswa pada SPJJ.
2. Membuat model keberhasilan belajar mahasiswa pada SPJJ dengan menggunakan model logit dan model probit.

## **B. LANDASAN TEORI**

### **1. Sistem Pembelajaran Jarak Jauh (SPJJ)**

SPJJ adalah proses transfer pengetahuan untuk pelajar (siswa) yang dipisahkan dari instruktur (guru) dengan waktu atau jarak fisik sehingga membuat penggunaan komponen teknologi, seperti video, internet, CD, kaset dan bentuk teknologi lainnya untuk mencapai pembelajaran. Dalam sistem pembelajaran jarak jauh (SPJJ), mahasiswa dapat mempelajari materi yang diberikan dan dipelajari diberbagai tempat seperti di rumah, tempat kerja, di kelas, di perpustakaan, atau bahkan ketika mereka melakukan perjalanan [5].

### **2. Karakteristik Sistem Pembelajaran Jarak Jauh (SPJJ)**

Universitas Terbuka adalah perguruan tinggi yang menggunakan SPJJ berbeda dibandingkan dengan Universitas Konvensional. Perbedaan yang menonjol adalah sistem pembelajarannya. Salah satu pendapat yang sering dikutip dan dijadikan acuan dalam banyak tulisan tentang pendidikan jarak jauh adalah pendapat Keegan. Keegan dalam [9] menyatakan bahwa ada 6 (enam) karakteristik utama dari sistem pembelajaran jarak jauh (SPJJ) ini, yaitu:

- a. Terpisahnya pengajar dengan pembelajar yang membedakan SPJJ dengan sistem tatap muka.
- b. Pengaruh suatu lembaga pendidikan merupakan hal yang membedakannya dengan studi pribadi.

- c. Menggunakan media teknis seperti media cetak, audio, video dan komputer untuk menyatukan pengajar dengan pembelajar dalam menyampaikan isi pendidikan.
- d. Disediaknya komunikasi dua arah sehingga pembelajar memperoleh keuntungan dan bahkan mahasiswa dapat berinisiatif untuk berdialog.
- e. Adanya kemungkinan sewaktu-waktu bertemu untuk keperluan belajar dan sosialisasi.
- f. Kehadiran SPJJ lebih membawa ke arah industrialisasi pendidikan.

### 3. Keberhasilan Belajar

Keberhasilan belajar identik dengan prestasi belajar yang dicapai mahasiswa setelah melakukan proses belajar. Dalam mencapai hasil yang optimal, mahasiswa diharapkan dapat belajar secara mandiri baik dilakukan secara sendiri maupun berkelompok [2].

Keberhasilan belajar dipengaruhi oleh beberapa faktor yaitu faktor internal dan faktor eksternal. Faktor internal meliputi keadaan fisik, intelegensi, minat, bakat, dan motivasi. Sedangkan faktor eksternal meliputi faktor lingkungan, faktor sekolah dan faktor keluarga [6].

### 4. Regresi Logistik Biner

Regresi logistik biner merupakan suatu metode regresi logistik yang digunakan untuk mencari hubungan antara variabel respon ( $y$ ) yang bersifat biner atau dikotomus dengan variabel prediktor ( $x$ ) yang bersifat polikotomus [3].

Model umum regresi logistik dengan  $p$  faktor (banyaknya variabel prediktor) adalah sebagai berikut:

$$\pi(x) = \frac{e^{(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)}}{1 + e^{(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)}} \quad (2.2)$$

Keterangan:

$\pi(x)$  : Peluang

$\beta_0, \beta_1 \dots \beta_p$  : Parameter

## 5. Regresi Probit

Model probit pertama kali diperkenalkan oleh Chester Bliss pada tahun 1934 dalam aplikasi di bidang teknologi [1]. Model probit merupakan model nonlinier yang menggunakan bilangan biner (variabel boneka) sebagai variabel responnya dan mengandaikan faktor galat  $\varepsilon_i$  berdistribusi Normal.

[7] menyatakan bahwa misalkan  $Y_i^*$  adalah variabel respon yang tak teramati. Variabel respon ini dipengaruhi oleh variabel prediktor sehingga persamaan ditulis dalam bentuk.

$$Y_i^* = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi} + \varepsilon_i \quad (2.10)$$

Model probit dapat dinyatakan dalam bentuk persamaan sebagai berikut:

$$P_i = \Phi(Y_i) = \Phi(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi}) \quad (2.11)$$

Dengan menggunakan invers fungsi  $\Phi$  maka persamaan (2.11) dapat dinyatakan dalam bentuk linier dengan persamaan sebagai berikut:

$$\Phi^{-1}(P_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi} \quad (2.12)$$

Keterangan:

$P_i$  : Peluang "Sukses" ke- $i$

$\Phi(Y_i)$  : Fungsi Distribusi Kumulatif Normal

## C. METODE PENELITIAN

### 1. Data Penelitian

#### a. Jenis dan Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder dan data primer. Data sekunder berupa data mahasiswa JME-UT yang sudah memiliki Indeks Prestasi Kumulatif (IPK). Data primer berupa data yang dikumpulkan dari kuesioner yang dibagikan kepada 172 mahasiswa JME-UT.

#### b. Variabel Penelitian

Variabel penelitian yang digunakan dalam penelitian ini adalah:

a) Variabel Respon (*dependent variable*)

Variabel respon yang digunakan dalam penelitian ini adalah Indeks Prestasi Kumulatif (IPK) mahasiswa JME-UT dengan kategori “Berhasil” apabila  $IPK \geq 3$  dan kategori “Tidak Berhasil” apabila  $IPK < 3$ .

b) Variabel Prediktor (*independent variable*)

Variabel prediktor yang digunakan dalam penelitian ini adalah:

- a. Umur ( $X_1$ )
- b. Jenis Kelamin ( $X_2$ )
- c. Status Pernikahan ( $X_3$ )
- d. Status Pekerjaan ( $X_4$ )
- e. Motivasi Belajar Mahasiswa ( $X_5$ )
- f. Fasilitas Selama Belajar ( $X_6$ )
- g. Waktu belajar diluar jam tutorial online dan kuliah tatap muka ( $X_7$ )
- h. Jumlah sks yang diambil ( $X_8$ )
- i. Keikutsertaan dalam belajar ( $X_9$ )
- j. Gaya Belajar mandiri ( $X_{10}$ )

## 2. Metode Analisis Penelitian

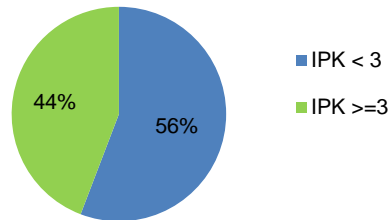
Tahapan analisis dalam penelitian ini adalah sebagai berikut:

1. Menguji variabel-variabel yang diduga berpengaruh dalam keberhasilan belajar mahasiswa pada SPJJ.
  - a. Melakukan uji serentak seluruh variabel prediktor terhadap variabel respon.
  - b. Melakukan uji parsial masing-masing variabel prediktor terhadap variabel respon.
2. Membuat model keberhasilan belajar mahasiswa pada SPJJ menggunakan model logit dan model probit.
  - a. Melakukan pembentukan model logit dan model probit.
  - b. Menguji kesesuaian model logit dan model probit yang telah terbentuk.
  - c. Melakukan ketepatan klasifikasi pada model logit dan model probit.
  - d. Melakukan uji pemilihan model terbaik dilihat dari nilai statistik Hosmer-Lemeshow, statistik Pearson dan statistik Devians.
  - e. Kesimpulan dari hasil analisis.

## D. HASIL DAN PEMBAHASAN

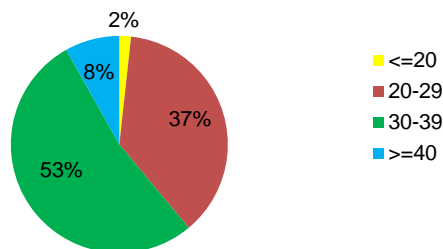
### 1. Deskripsi Data Keberhasilan Belajar Mahasiswa pada Sistem Pembelajaran Jarak Jauh (SPJJ)

Data dalam penelitian ini menggunakan data sekunder dan data primer. Terdapat dua macam variabel yang digunakan, yaitu variabel respon dan variabel prediktor. Variabel respon dalam penelitian yaitu Indeks Prestasi Kumulatif (IPK), sedangkan variabel prediktornya yaitu umur, jenis kelamin, status nikah, status kerja, waktu belajar diluar jam tutorial online dan kuliah tatap muka, jumlah sks yang diambil, keikutsertaan dalam belajar dan gaya belajar mandiri. Persentase jumlah untuk tiap kategori dari masing-masing variabel adalah sebagai berikut:



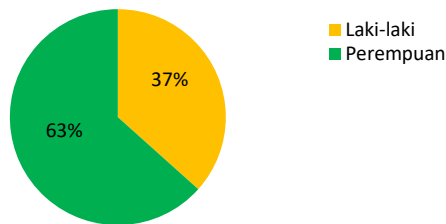
**Gambar 4.1. Persentase Indeks Prestasi Kumulatif (IPK) Mahasiswa JME-UT**

Berdasarkan Gambar 4.1 dapat dilihat bahwa persentase jumlah Indeks Prestasi Kumulatif (IPK) mahasiswa JME-UT dengan  $IPK < 3$  lebih besar dibandingkan  $IPK \geq 3$ . Mahasiswa dengan  $IPK < 3$  sebesar 56 % atau berjumlah 96 orang, sedangkan mahasiswa dengan  $IPK \geq 3$  sebesar 44 % atau berjumlah 76 orang.



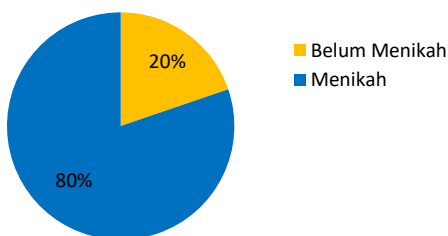
**Gambar 4.2. Persentase Usia Mahasiswa JME-UT**

Pada Gambar 4.2 dapat dilihat bahwa persentase jumlah usia mahasiswa JME-UT yang berusia 30 – 39 tahun sebesar 53 % atau berjumlah 91 orang, mahasiswa yang berusia 20 – 29 tahun sebesar 37 % atau berjumlah 64 orang, mahasiswa yang berusia  $\geq 40$  tahun sebesar 8 % atau berjumlah 14 orang, dan mahasiswa yang berusia  $\leq 20$  tahun sebesar 2 % atau berjumlah 3 orang.



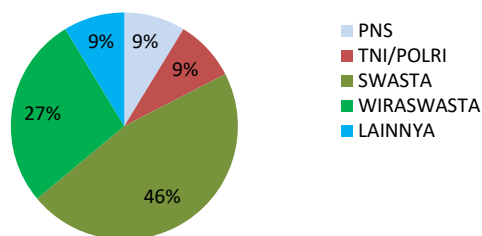
**Gambar 4.3. Persentase Jenis Kelamin Mahasiswa JME-UT**

Perhatikan Gambar 4.3 dapat dilihat bahwa persentase jumlah jenis kelamin mahasiswa JME-UT yang berjenis kelamin Perempuan sebesar 63 % atau berjumlah 109 orang, sedangkan mahasiswa yang berjenis kelamin Laki-laki sebesar 37 % atau berjumlah 63 orang.

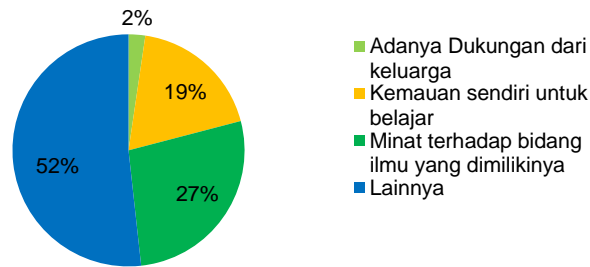


**Gambar 4.4. Persentase Status Nikah Mahasiswa JME-UT**

Gambar 4.4 dapat dilihat bahwa persentase jumlah status nikah mahasiswa JME-UT yang sudah menikah sebesar 80 % atau berjumlah 138 orang, sedangkan yang belum menikah sebesar 20 % atau berjumlah 34 orang.

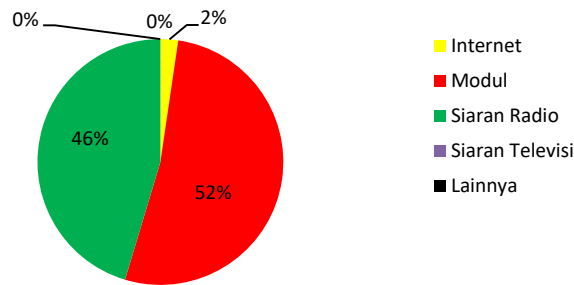


**Gambar 4.5. Persentase Status Kerja Mahasiswa JME-UT**



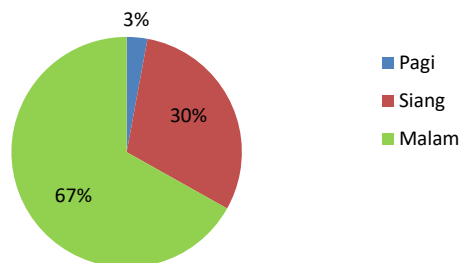
**Gambar 4.6. Persentase Motivasi Belajar Mahasiswa JME-UT**

Pada Gambar 4.6 dapat dilihat bahwa persentase jumlah motivasi mahasiswa JME-UT dengan kategori lainnya dari yang telah disebutkan sebesar 52 % atau berjumlah 89 orang, minat terhadap bidang ilmu yang dimilikinya sebesar 27 % atau berjumlah 47 orang, kemauan sendiri untuk belajar sebesar 19 % atau berjumlah 32 orang, dan adanya dukungan dari keluarga sebesar 2 % atau berjumlah 4 orang.



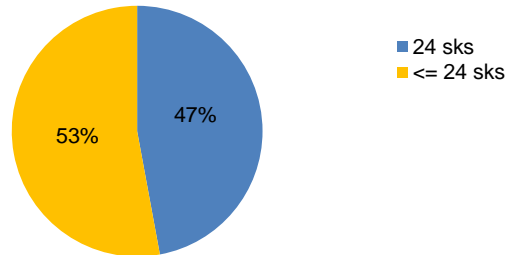
**Gambar 4.7. Persentase Fasilitas Selama Belajar Mahasiswa JME-UT**

Berdasarkan Gambar 4.7 dapat dilihat bahwa persentase fasilitas selama belajar JME-UT yang menggunakan modul sebesar 52 % atau sebanyak 90 orang, siaran radio sebesar 46 % atau berjumlah 78 orang, internet sebesar 2 % atau berjumlah 4 orang, siaran televisi dan lainnya sebesar 0 % artinya tidak ada yang menggunakan fasilitas siaran televisi serta fasilitas selain yang telah disebutkan.



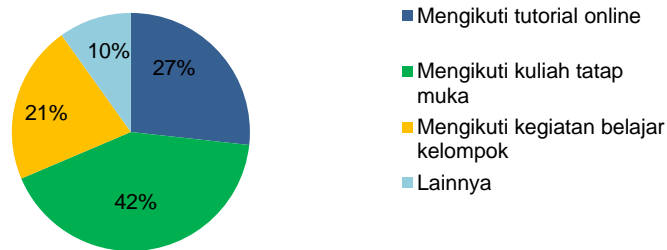
**Gambar 4.8. Persentase Waktu Belajar Mahasiswa JME-UT**

Gambar 4.8 dapat dilihat bahwa persentase jumlah waktu belajar di luar jam tutorial online dan tatap muka mahasiswa JME-UT dengan kategori belajar di malam hari sebesar 67 % atau berjumlah 115 orang, belajar di siang hari sebesar 30 % atau berjumlah 52 orang, dan belajar di pagi hari sebesar 3 % atau berjumlah 5 orang.



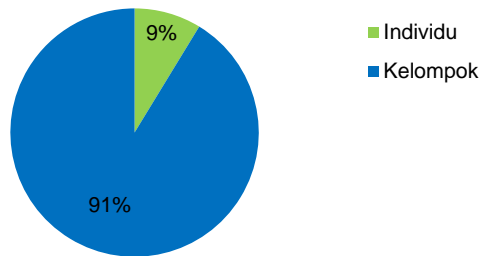
**Gambar 4.9. Persentase Jumlah sks Mahasiswa JME-UT**

Pada Gambar 4.9 dapat dilihat bahwa persentase jumlah sks mahasiswa JME-UT dengan kategori jumlah sks kurang dari 24 sebesar 53 % atau berjumlah 91 orang, sedangkan mahasiswa dengan jumlah sks sama dengan 24 sebesar 47 % atau berjumlah 81 orang.



**Gambar 4.10. Persentase Keikutsertaan dalam Belajar Mahasiswa JME-UT**

Berdasarkan Gambar 4.10 dan Tabel 4.10 dapat dilihat bahwa persentase jumlah keikutsertaan dalam belajar mahasiswa JME-UT dengan kategori yang mengikuti kuliah tatap muka sebesar 42 % atau berjumlah 72 orang, mengikuti tutorial online sebesar 27 % atau berjumlah 46 orang, mengikuti kegiatan belajar kelompok sebesar 21 % atau berjumlah 37 orang dan lainnya sebanyak 10 % atau sekitar 17 orang.



**Gambar 4.11. Persentase Gaya Belajar Mandiri Mahasiswa JME-UT**

Pada Gambar 4.11 dapat dilihat bahwa persentase jumlah gaya belajar mandiri mahasiswa JME-UT dengan kategori gaya belajar kelompok sebesar 91 % atau berjumlah 157 orang, sedangkan gaya belajar individu sebesar 9 % atau berjumlah 15 orang.

## 2. Uji Validitas dan Uji Reliabilitas

### a. Uji Validitas

Hasil yang diperoleh dari program SPSS 21 untuk menguji validitas:

**Tabel 4.12. Hasil Uji Validitas**

Variabel	Nama Variabel	<i>Pearson Correlation</i>	Keterangan
$X_1$	Umur	0,316**	Valid
$X_2$	Status Nikah	0,263**	Valid
$X_3$	Status Kerja	0,599**	Valid
$X_4$	Jenis Kelamin	0,121	Tidak Valid
$X_5$	Motivasi dalam belajar	0,386**	Valid
$X_6$	Fasilitas selama belajar	0,259**	Valid
$X_7$	Waktu belajar diluar jam tutorial online dan tatap muka	0,097	Tidak Valid
$X_8$	Jumlah sks yang diambil	0,295**	Valid
$X_9$	Keikutsertaan dalam belajar	0,500**	Valid
$X_{10}$	Gaya belajar mandiri	0,205**	Valid

Berdasarkan Tabel 4.12 dapat dilihat bahwa pertanyaan 1, 2, 3, 5, 6, 8, 9, dan 10 valid karena nilai *Pearson Correlation* > r-tabel yaitu  $X_1 = 0,360$ ,  $X_2 = 0,263$ ,  $X_3 = 0,599$ ,  $X_5 = 0,386$ ,  $X_6 = 0,259$ ,  $X_8 = 0,295$ ,  $X_9 = 0,500$ ,  $X_{10} = 0,205$  sedangkan pertanyaan 4 dan 7 tidak valid karena nilai *Pearson Correlation* < r-tabel yaitu  $X_4 = 0,121$ ,  $X_7 = 0,097$ . Artinya ada 8 pertanyaan yang dapat digunakan dalam kuesioner. Selanjutnya akan dilakukan uji validitas ulang dengan membuang butir pertanyaan yang tidak valid.

**Tabel 4.13. Hasil uji validitas tahap dua**

Variabel	Nama Variabel	<i>Pearson Correlation</i>	Keterangan
$X_1$	Umur	0,360**	Valid
$X_2$	Status Nikah	0,314**	Valid
$X_3$	Status Kerja	0,616**	Valid
$X_5$	Motivasi dalam belajar	0,434**	Valid
$X_6$	Fasilitas selama belajar	0,312**	Valid
$X_8$	Jumlah SKS yang diambil	0,278**	Valid
$X_9$	Keikutsertaan dalam belajar	0,490**	Valid
$X_{10}$	Gaya belajar mandiri	0,203**	Valid

Hasil uji validitas pada Tabel 4.13 dapat dilihat bahwa butir pertanyaan 1, 2, 3, 5, 6, 8, 9, dan 10 sudah valid terlihat dari nilai *Pearson Correlation* > r-tabel. Maka butir pertanyaan tersebut dapat digunakan dalam kuesioner dan dapat digunakan sebagai variabel prediktor untuk dilakukan analisis. Selanjutnya akan dilakukan uji reliabilitas untuk melihat apakah kuesioner yang baru reliabel atau tidak. Suatu kuesioner dapat dikatakan reliabel apabila nilai *Cronbach' Alpha* lebih besar dari nilai r-tabel (r-tabel = 0,126). Uji reliabilitas juga dilakukan dengan SPSS 21.

**Tabel 4.14. Hasil Uji Realibilitas**

<i>Cronbach's Alpha</i>	<i>N of Items</i>
0,219	8

Pada Tabel 4.14 diketahui bahwa nilai *Cronbach's Alpha* lebih besar dari r-tabel, yaitu  $0,219 > 0,126$ , artinya kuesioner dengan 8 butir pertanyaan tersebut sudah reliabel dan dapat digunakan untuk analisis.

### 3. Hasil Analisis Regresi Logistik Biner

#### a. Uji Signifikansi Parameter Model Logit secara Serentak (Simultan)

Hasil pengujian serentak dengan menggunakan program SPSS 21 dapat dilihat pada tabel 4.15.

**Tabel 4.15. Hasil Likelihood Ratio Test**

	Chi-square	df	Sig.
Step	118.548	8	.000
Step 1 Block	118.548	8	.000
Model	118.548	8	.000

Berdasarkan Tabel 4.15, diketahui bahwa nilai statistik uji *likelihood ratio* ( $G$ ) sebesar 118,548 dan  $p$  – *value* sebesar 0,000. Pengambilan keputusan dilakukan dengan membandingkan nilai  $p$  – *value* dari  $G$  sebesar 0,000 lebih kecil dari 0,10 ( $\alpha$ ), sehingga  $H_0$  ditolak. Artinya minimal ada satu variabel prediktor yang berpengaruh secara simultan terhadap variabel respon.

**b. Uji Signifikansi Parameter Model Logit secara Parsial (Uji Wald)**

Hasil pengujian parsial dengan menggunakan program SPSS 21 dapat dilihat pada tabel 4.16.

**Tabel 4.16. Hasil Uji Parsial**

	B	S.E.	Wald	df	Sig.	Exp(B)
Umur	-.621	.412	2.279	1	.131	.537
StatusNikah	.856	.645	1.761	1	.184	2.353
StatusKerja	-.032	.248	.016	1	.898	.969
Motivasi	.564	.301	3.512	1	.061	1.758
Step 1 <sup>a</sup> Fasilitas	.221	.450	.243	1	.622	1.248
JumlahSKS	4.253	.552	59.415	1	.000	70.344
KeikutsertaanBelajar	.042	.268	.024	1	.876	1.043
GayaBelajar	-.085	1.009	.007	1	.933	.919
Constant	-8.116	2.961	7.516	1	.006	.000

a. Variable(s) entered on step 1: Umur, StatusNikah, StatusKerja, Motivasi, Fasilitas, JumlahSKS, KeikutsertaanBelajar, GayaBelajar.

Berdasarkan penjelasan sebelumnya, variabel yang layak masuk dalam model regresi logistik adalah variabel motivasi belajar mahasiswa ( $X_5$ ) dan variabel jumlah sks ( $X_8$ ) dalam kasus faktor-faktor yang mempengaruhi keberhasilan belajar mahasiswa pada SPJJ.

**c. Pembentukan Model Logit**

Berdasarkan tabel 4.16 pada kolom  $B$  dan pengujian signifikansi parameter diperoleh model logit untuk faktor-faktor yang mempengaruhi keberhasilan belajar mahasiswa pada SPJJ dengan persamaan sebagai berikut:

$$\pi(x) = \frac{e^{(-8,116+0,564X_5+4,253X_8)}}{1 + e^{(-8,116+0,564X_5+4,253X_8)}}$$

#### d. Uji Kesesuaian Model untuk Model Logit

Uji kesesuaian model regresi logistik dengan menggunakan uji Hosmer dan Lemeshow, uji Pearson, dan uji Devians untuk mengetahui apakah model yang digunakan telah sesuai atau tidak. Hipotesis yang diuji adalah sebagai berikut:

$H_0$  : Model sesuai (tidak terdapat perbedaan yang signifikan antara hasil pengamatan dengan kemungkinan hasil prediksi model)

$H_1$  : Model tidak sesuai (terdapat perbedaan yang signifikan antara hasil pengamatan dengan kemungkinan hasil prediksi model)

**Tabel 4.17. Hasil Pearson test, Devians test, Hosmer-Lemeshow test**

Method	Chi-Square	DF	P
Pearson	125.998	136	0.719
Deviance	93.151	136	0.998
Hosmer-Lemeshow	7.366	8	0.498

#### 4. Hasil Analisis Regresi Probit

##### a. Uji Signifikansi Parameter Model Probit secara Serentak (Simultan)

Berdasarkan Lampiran 10 diketahui bahwa nilai statistik uji *likelihood ratio* ( $G$ ) sebesar 119,128 dan  $p$  – *value* sebesar 0,000. Pengambilan keputusan dilakukan dengan membandingkan nilai  $p$  – *value* dari  $G$  sebesar 0,000 lebih kecil dari 0,10 ( $\alpha$ ), sehingga  $H_0$  ditolak. Artinya minimal ada satu variabel prediktor yang berpengaruh secara simultan terhadap variabel respon.

##### b. Uji Signifikansi Parameter Model Probit secara Parsial

Setelah dilakukan uji signifikansi parameter secara simultan, selanjutnya dilakukan pengujian parsial dengan membandingkan dugaan  $\beta$  dengan penduga standar errornya.

Berdasarkan penjelasan sebelumnya, variabel yang layak masuk dalam model regresi logistik adalah variabel umur ( $X_1$ ), variabel motivasi belajar mahasiswa ( $X_5$ ) dan variabel jumlah sks ( $X_8$ ) dalam kasus faktor-faktor yang mempengaruhi keberhasilan belajar mahasiswa pada SPJJ.

**c. Pembentukan Model Probit**

Berdasarkan lampiran 10 pada kolom *P* dan pengujian signifikansi parameter diperoleh model probit untuk faktor-faktor yang mempengaruhi keberhasilan belajar mahasiswa pada SPJJ dengan persamaan sebagai berikut:

$$P_i = \Phi(Y) = -4,681 - 0,359X_{1i} + 0,327X_{5i} + 2,474X_{8i}$$

**d. Uji Kesesuaian Model untuk Model Probit**

Uji kesesuaian model regresi logistik dengan menggunakan uji Hosmer dan Lemeshow, uji Pearson, dan uji Devians untuk mengetahui apakah model yang digunakan telah sesuai atau tidak. Hipotesis yang diuji adalah sebagai berikut:

$H_0$  : Model sesuai (tidak terdapat perbedaan yang signifikan antara hasil pengamatan dengan kemungkinan hasil prediksi model)

$H_1$  : Model tidak sesuai (terdapat perbedaan yang signifikan antara hasil pengamatan dengan kemungkinan hasil prediksi model)

**Tabel 4.18. Hasil Pearson test, Devians test, Hosmer-Lemeshow test**

Method	Chi-Square	DF	P
Pearson	126.037	136	0.719
Deviance	92.572	136	0.998
Hosmer-Lemeshow	6.415	8	0.601

**5. Ketepatan Klasifikasi Model Logit dan Model Probit**

**a. Ketepatan Klasifikasi Model Logit**

Hasil ketepatan klasifikasi hasil prediksi model logit dapat diketahui dengan menggunakan Tabel 4.19 sebagai berikut:

**Tabel 4.19. Ketepatan Hasil Prediksi Model Logit**

Observed			Predicted		
			IPK		Percentage Correct
			>= 3	< 3	
Step 1	IPK	>= 3	68	8	89.5
		< 3	13	83	86.5
	Overall Percentage				87.8

## b. Ketepatan Klasifikasi Model Probit

Hasil ketepatan klasifikasi hasil prediksi model probit dapat diketahui dengan menggunakan Tabel 4.20 sebagai berikut:

**Tabel 4.20. Ketepatan Hasil Prediksi Model probit**

Observed			Predicted		
			IPK		Percentage Correct
			$\geq 3$	$< 3$	
Step 1	IPK	$\geq 3$	71	5	93.4
		$< 3$	26	70	72,9
	Overall Percentage				81.9

## 6. Pemilihan Model Terbaik

Pemilihan model terbaik didapat dari model probit karena nilai *Chi-Square* Devians dan nilai *Chi-Square* Hosmer-Lemeshow lebih kecil dibanding nilai *Chi-Square* Devians dan nilai *Chi-Square* Hosmer-Lemeshow model logit, dimana nilai *Chi-Square* Devians dan nilai *Chi-Square* Hosmer-Lemeshow model probit berturut-turut, yaitu 92,572 dan 6,415, sedangkan nilai *Chi-Square* Devians dan nilai *Chi-Square* Hosmer-Lemeshow model logit berturut-turut, yaitu 93,151 dan 7,366. Hasil perhitungan nilai *Chi-Square* Pearson model logit lebih baik dari model probit, karena nilai *Chi-Square* pearson model logit lebih kecil dari model probit yaitu  $125,998 < 126,037$ .

## E. KESIMPULAN

### 1. Kesimpulan

Berdasarkan hasil dan pembahasan yang telah diuraikan pada bab sebelumnya, diperoleh kesimpulan sebagai berikut:

- Pada model logit faktor yang berpengaruh dalam keberhasilan belajar mahasiswa pada sistem pembelajaran jarak jauh (SPJJ) adalah motivasi belajar mahasiswa dan jumlah sks. Sedangkan pada model probit faktor yang berpengaruh dalam keberhasilan belajar mahasiswa pada sistem pembelajaran jarak jauh (SPJJ) adalah umur, motivasi belajar mahasiswa dan jumlah sks.
- Model keberhasilan belajar mahasiswa pada SPJJ dengan persamaan yang diperoleh adalah sebagai berikut:

Model Logit : 
$$\pi(x) = \frac{e^{(-8,116+0,564X_5+4,253X_8)}}{1 + e^{(-8,116+0,564X_5+4,253X_8)}}$$

Model Probit : 
$$P_i = \Phi(Y) = -4,681 - 0,359X_{1i} + 0,327X_{5i} + 2,474X_{8i}$$

## 2. Saran

Penelitian selanjutnya dapat menggunakan faktor-faktor yang mempengaruhi keberhasilan belajar mahasiswa pada SPJJ yang belum digunakan peneliti dalam kasus ini.

### DAFTAR PUSTAKA

- [1] Agresti, A., 2007, *Categorical Data Analysis*, New York: John Wiley and Sons.
- [2] Anonim, 2016, *Katalog Sistem Penyelenggaraan FEKON, FISIP, FMIPA, FKIP Universitas Terbuka*. Tangerang Selatan: Universitas Terbuka.
- [3] Hosmer, D.W. and S., Lemeshow, 2000, *Applied Logistic Regression, Second Edition*, USA: John Wiley and Sons.
- [4] Isfarudi, 1994, *Faktor-faktor Penentu Resistensi Belajar Mahasiswa FMIPA Universitas Terbuka*, Program Pasca Sarjana. Institut Keguruan dan Ilmu Pendidikan, Jakarta, Tesis.
- [5] Moore, M.G., dan G., Kearsley, 1996, *Distance Education. A Systems View*. Belmont, CA: Wadsworth Publishing Company.
- [6] Mutijah, 2015, *Analisis Keberhasilan Belajar Mahasiswa Penerima Beasiswa di IAIN Purwokerto Menggunakan Regresi Logistik*, Laporan Penelitian, IAIN Purwokerto.
- [7] Nagler, J . 1994. *Interpreting Probit Analysis*. Diakses tanggal 28 Februari 2017. <http://www.nyu.edu/classes/nagler/quant2/notes/probit1.pdf>.
- [8] Nikmah, N., 2014, *Faktor-faktor yang Mempengaruhi Nilai Indeks Prestasi Kumulatif (IPK) Mahasiswa Program Studi D-III Kebidanan STIKES INSAN SE AGUNG BANGKALAN*, STIKES Insan Se Agung Bangkalan.
- [9] Rumble, G., 1989, *On Defining Distance Education*, The American Journal Of Distance Education, 3. (2), 8-21.
- [10] Yong, B., 2003, *Penaksiran Maksimum Likelihood bagi Model Probit dan Model Probit Bivariat*. Diakses tanggal 28 Februari 2017. <http://home.unpar.ac.id/~integral/Volume%208/integral%201%20%201/penaksiran%20maksimum%likelihoood.pdf>.

# MODEL LOG LINIER PADA TABEL KONTINGENSI TAK SEMPURNA BERDIMENSI EMPAT (Studi Kasus: Data Jumlah Pegawai Negeri Sipil Provinsi Jawa Barat Tahun 2012)

Dwi Puspa Oktarina<sup>1</sup>, Sigit Nugroho<sup>2</sup> dan Etis Sunandi<sup>2</sup>

<sup>1</sup>Mahasiswa Jurusan Statistika, Prodi Matematika, FMIPA, UNIB, Bengkulu

<sup>2</sup>Dosen Pembimbing Jurusan Statistika, Prodi Matematika, FMIPA, UNIB, Bengkulu

[Dwipuspa002@gmail.com](mailto:Dwipuspa002@gmail.com)

## ABSTRAK

Tujuan dalam penulisan ini adalah untuk mengetahui prosedur analisis model log linier pada tabel kontingensi tak sempurna berdimensi empat dan menerapkannya pada data Jumlah Pegawai Negeri Sipil Provinsi Jawa Barat Tahun 2012. Data kategorik merupakan data hasil pengamatan sampel dalam suatu populasi yang memiliki keadaan serupa yang dikelompokkan silang kedalam beberapa variabel kategorik. Suatu tabel kontingensi dikatakan tak sempurna, jika dan hanya jika tabel tersebut mempunyai sebuah sel kosong atau lebih untuk populasi yang ditinjau. Analisis data kategorik dapat dilakukan dengan menggunakan model *log linier* untuk menganalisis hubungan antara variabel-variabel kategorik yang membentuk tabel kontingensi sembarang dimensi. Model *log linier* empat dimensi merupakan perluasan dari model *log linier* tiga dimensi. Model *log linier* pada tabel kontingensi tak sempurna adalah model *log linier* yang dilakukan pada tabel yang mempunyai sebuah sel kosong atau lebih untuk populasi yang ditinjau dengan dimensi empat. Bentuk umum model *log linier* empat dimensi yaitu  $\log m_{ijkl} = \mu + \lambda_i^W + \lambda_j^X + \lambda_k^Y + \lambda_l^Z + \lambda_{ij}^{WX} + \lambda_{ik}^{WY} + \lambda_{il}^{WX} + \lambda_{jk}^{XY} + \lambda_{jl}^{XZ} + \lambda_{kl}^{YZ} + \lambda_{ijk}^{WXY} + \lambda_{ijl}^{WXZ} + \lambda_{jkl}^{XYZ} + \lambda_{ijkl}^{WXYZ}$ . Berdasarkan analisis data yang dilakukan model terbaik yang terpilih adalah model dua arah yakni  $\log m_{ijkl} = \mu + \lambda_i^W + \lambda_j^X + \lambda_k^Y + \lambda_l^Z + \lambda_{ij}^{WX} + \lambda_{ik}^{WY} + \lambda_{il}^{WX} + \lambda_{jk}^{XY} + \lambda_{jl}^{XZ} + \lambda_{kl}^{YZ}$ .

**Kata Kunci:** Data Kategorik, tabel kontingensi, tabel kontingensi tak sempurna, model log linier

## 1. Pendahuluan

Dalam kehidupan sehari-hari sering ditemukan data yang dikelompokkan ke dalam suatu kategori tertentu. Misalkan saja pada bidang kesehatan, pendidikan, ekonomi dan lain-lain (Hapsari, 2011). Menurut Lestyorini (2010), data yang terdiri dari beberapa kategorik ini disebut data kategorik, contohnya jenis pekerjaan yang terbagi menjadi: pegawai negeri dan pegawai swasta. Sehingga data kategorik merupakan data hasil pengamatan sampel dalam suatu populasi yang memiliki keadaan serupa yang dikelompokkan silang kedalam beberapa variabel kategorik (Fienberg, 2007).

Analisis dari data kategorik diaplikasikan dalam sebuah tabel yang menggambarkan frekuensi dari pengamatan yang terjadi pada level keragaman kombinasi dari variabel. Tabel yang mengaplikasikan data kategorik disebut dengan tabel kontingensi (Agresti, 2007).

Tabel kotingensi atau yang sering disebut tabulasi silang (*cross tabulation* atau *cross classification*) adalah tabel yang berisi data jumlah atau frekuensi atau beberapa klasifikasi (kategori). Metode *cross tabulation* dapat menjawab hubungan antara dua, tiga atau lebih variabel penelitian tetapi bukan hubungan sebab akibat. Semakin banyak jumlah variabel yang ditabulasikan maka semakin baik interpretasinya (Agresti, 2007).

Tabel kontingensi pada umumnya berbentuk tabel sempurna, namun ada juga tabel kontingensi tidak sempurna. Suatu tabel dikatakan tak sempurna, jika dan hanya jika tabel tersebut mempunyai sebuah sel kosong atau lebih untuk populasi yang ditinjau (Hapsari, 2011).

Analisis data kategorik dapat dilakukan dengan menggunakan model log linier. Model log linier digunakan untuk menganalisis hubungan antara variabel-variabel kategorik yang membentuk tabel kontingensi sembarang dimensi (Lestyorini, 2010).

Beberapa penelitian yang telah membahas tentang model log linier yang diterapkan pada data hipertensi, model relational, model log linier pada data korupsi, model graph QS, model hierarchichal dan model log linier pada kecelakaan lalu lintas. Oleh karena itu penulis tertarik melakukan penelitian model log linier pada tabel kontingensi berdimensi 4 dan menerapkannya dalam kasus jumlah pegawai negeri sipil Provinsi Jawa Barat tahun 2012 berdasarkan unit kerja/institusi, tingkat pendidikan formal, jenis kelamin dan golongan, dalam kategori tertentu terdapat sel kosong karena tidak ada yang memenuhi kategori tersebut. Penelitian yang menggunakan analisis model log linier pada tabel kontingensi tak sempurna berdimensi empat ini dilakukan untuk mengetahui variabel-variabel yang saling berhubungan antara keempat variabel yang diamati.

## 2. Prosedur Model log Linier Pada Tabel Kontingensi Sempurna

*Log linier* merupakan suatu metode yang digunakan untuk mengetahui sel-sel mana yang mempengaruhi dan menyebabkan hubungan antar variabel bersifat dependen. Model *log linier* adalah suatu model untuk memperoleh model statistika yang menyatakan hubungan antara variabel dengan data yang bersifat kualitatif (skala nominal atau ordinal) (Wulandari, Salamah dan Susilaningrum, 2009).

Tabel kontingensi atau yang sering disebut tabulasi silang (cross tabulation atau cross classification) adalah tabel yang berisi data jumlah atau frekuensi atau beberapa klasifikasi (kategori). Metode tabel kontingensi dapat menjawab hubungan dua atau lebih variabel penelitian tetapi bukan hubungan sebab akibat, semakin banyak jumlah variabel yang ditabulasikan maka semakin baik interpretasinya (Wulandari, Salamah dan Susilaningrum, 2009).

Menurut Agresti (2007), data bivariat (X,Y) dengan X dan Y merupakan variabel kategorik, yang masing-masing mempunyai i baris dan j kolom. Data bivariat dapat disajikan dalam bentuk tabel frekuensi menurut variabel X dan Y yang mempunyai i baris dan j kolom. Tabel seperti ini disebut tabel kontingensi  $i \times j$  yang secara umum dapat disajikan dalam Tabel 2.1 di bawah ini (Wulandari, Salamah dan Susilaningrum, 2009).

Tabel 1. Tabel kontingensi  $i \times j$

		Variabel 2 (Y)				Total
		$Y_1$	$Y_2$	...	$Y_j$	
Variabel (X)	$X_1$	$n_{11}$	$n_{12}$	...	$n_{1j}$	$n_{1+}$
	$X_2$	$n_{21}$	$n_{22}$	...	$n_{2j}$	$n_{2+}$
	⋮	⋮	⋮	⋮	⋮	⋮
	$X_i$	$n_{i1}$	$n_{i2}$	...	$n_{ij}$	$n_{i+}$
Total		$n_{+1}$	$n_{+2}$	...	$n_{+j}$	$n_{++}$

Model *log linier independent* pada tabel kontingensi dua dimensi (Agresti, 2002):

$$\log m_{ij} = \mu + \lambda_i^x + \lambda_j^y; \text{ untuk } i = 1 \dots I, j = 1 \dots J \quad (1)$$

Keterangan:

- $\mu$  : efek rata-rata secara umum
- $\lambda_i^x$  : efek utama kategori ke- $i$  variabel X
- $\lambda_j^y$  : efek utama kategori ke- $j$  variabel Y

Dengan derajat bebasnya adalah

$$\begin{aligned} db &= db(\log m_{ij}) - (db(\mu) + db(\lambda_i^x) + db(\lambda_j^y)) \\ &= IJ - I - J + 1 \end{aligned}$$

Model di atas disebut sebagai model *log linier dependent* pada tabel kontingensi dua dimensi (Agresti, 2002)

$$\log m_{ij} = \mu + \lambda_i^x + \lambda_j^y + \lambda_{ij}^{xy}; \text{ untuk semua } i = 1 \dots I \text{ dan } j = 1 \dots J \quad (2)$$

Keterangan:

- $\mu$  : efek rata-rata secara umum
- $\lambda_i^x$  : efek utama kategori ke- $i$  variabel X
- $\lambda_j^y$  : efek utama kategori ke- $j$  variabel Y
- $\lambda_{ij}^{xy}$  : Parameter pengaruh variabel pertama dan kedua

### Prosedur Analisis Model Log Linier Dua Dimensi

Menurut Lestiyorini (2010), prosedur analisis pada model log linier dua dimensi adalah menggunakan uji independensi.

- Hipotesis
  - $H_0 : n_{ij} = n_{i+} \cdot n_{+j}$  (kedua variabel saling bebas)
  - $H_1 : n_{ij} \neq n_{i+} \cdot n_{+j}$  (kedua variabel tidak saling bebas)

- Taraf signifikansi  
 $\alpha = 0.05$

- Statistik Uji :

$\chi^2$

$$= \sum_{ij} \left[ \frac{(n_{ij} - m_{ij})^2}{m_{ij}} \right]$$

- Kriteria penolakan

Tolak  $H_0$  jika  $\chi_{hit}^2 > \chi_{tabel}^2$  atau  $p\_value \leq \alpha$

- Kesimpulan

jika  $\chi_{hit}^2 > \chi_{tabel}^2$  atau  $p\_value \leq \alpha$  maka  $H_0$  ditolak artinya  $n_{ij} \neq n_{i+} \cdot n_{+j}$  (kedua variabel tidak saling bebas) atau jika  $\chi_{hit}^2 \leq \chi_{tabel}^2$  atau  $p\_value > \alpha$  maka  $H_0$  diterima artinya  $n_{ij} = n_{i+} \cdot n_{+j}$  (kedua variabel saling bebas).

Tabel tiga dimensi mempunyai konsep yang sama dengan tabel kontingensi dua dimensi, hanya saja variabel yang digunakan lebih banyak yakni variabel X, Y dan Z. Menurut Wulandari, Salamah dan Susilaningrum (2009), tabel kontingensi tiga dimensi yang mempunyai ( $i \times j \times k$ ) sel, yang terdiri dari i baris, j kolom dan k layer yang disebut sebagai tabel kontingensi  $i \times j \times k$ . Tabel kontingensi  $i \times j \times k$  dapat dilihat pada Tabel 2.4 berikut:

Tabel 2. Tabel kontingensi  $i \times j \times k$

Variabel 1 (X)	Variabel 2 (Y)	Variabel 3 (Z)				Total
		$Z_1$	$Z_2$	...	$Z_k$	
$X_1$	$Y_1$	$n_{111}$	$n_{112}$	...	$n_{11k}$	$n_{11+}$
	⋮	⋮	⋮	...	⋮	⋮
	$Y_j$	$n_{1j1}$	$n_{1j2}$	...	$n_{1jk}$	$n_{1j+}$

⋮	⋮	⋮	⋮	⋮	⋮	⋮
$X_i$	$Y_1$	$n_{i11}$	$n_{i12}$	⋯	$n_{i1k}$	$n_{i1+}$
	⋮	⋮	⋮	⋯	⋮	⋮
	$Y_j$	$n_{ij1}$	$n_{ij2}$	⋯	$n_{ijk}$	$n_{ij+}$
<i>Total</i>		$n_{++1}$	$n_{++2}$	⋯	$n_{++k}$	$n_{+++}$

Model *independent* untuk tiga variabel adalah (Wulandari, Salamah dan Susilaningrum, 2009):

$$\log m_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z \quad (3);$$

$$i = 1 \dots I, j = 1 \dots J, k = 1 \dots K$$

Dimana :

$\mu$  : logaritma jumlah nilai harapannya atau rata-rata dari seluruh logaritma nilai harapannya.

$$\mu = \frac{\sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K \log m_{ijk}}{IJK} \quad (4)$$

$\lambda_i^X$  : pengaruh dari variabel 1 terhadap model.

$$\lambda_i^X = \frac{\sum_{j=1}^J \sum_{k=1}^K \log m_{ijk}}{JK} - \mu \quad (5)$$

$\lambda_j^Y$  : pengaruh dari variabel 2 terhadap model.

$$\lambda_j^Y = \frac{\sum_{i=1}^I \sum_{k=1}^K \log m_{ijk}}{IK} - \mu \quad (6)$$

$\lambda_k^Z$  : pengaruh dari variabel 3 terhadap model.

$$\lambda_k^Z = \frac{\sum_{i=1}^I \sum_{j=1}^J \log m_{ijk}}{IJ} - \mu \quad (7)$$

Dengan derajat bebasnya adalah

$$\begin{aligned} db &= db(\log m_{ijk}) - (db(\mu) + db(\lambda_i^X) + (\lambda_j^Y) + (\lambda_k^Z)) \\ &= IJK - ((1) + (I - 1) + (J - 1) + (K - 1)) \\ &= IJK - I - J - K + 2 \end{aligned}$$

Berdasarkan Agresti (2002), jika terdapat interaksi pada ketiga variabel, maka model menjadi

$$\log m_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ} + \lambda_{ijk}^{XYZ} \quad (8)$$

Keterangan:

$\mu$  : pengaruh rata-rata

$\lambda_i^X$  : pengaruh dari variabel 1 terhadap model.

$\lambda_j^Y$  : pengaruh dari variabel 2 terhadap model.

$\lambda_k^Z$  : pengaruh dari variabel 3 terhadap model.

$\lambda_l^W$  : pengaruh dari variabel 4 terhadap model.

$\lambda_{ij}^{XY}$  : pengaruh interaksi dari variabel 1 dan variabel ke-2

$\lambda_{ik}^{XZ}$  : pengaruh interaksi dari variabel 1 dan variabel ke-3

$\lambda_{jk}^{YZ}$  : pengaruh interaksi dari variabel 2 dan variabel ke-3

$\lambda_{ijk}^{XYZ}$  : pengaruh interaksi dari variabel 1, variabel 2 dan variabel ke-3

## Prosedur Analisis Model Log Linier Tiga Dimensi

Menurut Hapsari (2011), statistik cukup minimal untuk model-model *log linier* merupakan koefisien dari masing-masing parameternya yang diperoleh dari pengumpulan dan penjumlahan batas marjinal dari masing-masing parameternya.

Tabel 3. Statistik Cukup Minimal untuk Model *Log Linier Tiga Dimensi*

Simbol Model	Statistik Cukup Minimal
(XYZ)	$\{n_{ijk}\}$
(XY,Z)	$\{n_{ij+}, \{n_{++k}\}$
(XY,YZ)	$\{n_{ij+}, \{n_{+jk}\}$
(XZ,Y)	$\{n_{i+k}, \{n_{+j+}\}$
(YZ,X)	$\{n_{+jk}, \{n_{i++}\}$
(XZ,YZ)	$\{n_{i+k}, \{n_{+jk}\}$
(XY,XZ)	$\{n_{ij+}, \{n_{i+k}\}$
(XY,XZ,YZ)	$\{n_{ij+}, \{n_{i+k}, \{n_{+jk}\}$
(X,Y,Z)	$\{n_{i++}, \{n_{+j+}, \{n_{++k}\}$

(Hapsari, 2011)

Uji *goodness of fit* pada tabel kontingensi dilakukan untuk melihat apakah ada perbedaan antara frekuensi observasi dengan frekuensi yang diharapkan, langkah-langkah pengujiannya sebagai berikut (Agresti, 2007):

- Hipotesis  
 $H_0$  : Model sesuai dengan keadaan sebenarnya  
 $H_1$  : Model tidak sesuai dengan keadaan sebenarnya
- Taraf signifikansi  
 $\alpha = 0.05$
- Statistik Uji :  

$$\chi^2 = \sum_{ijk} \left[ \frac{(n_{ijk} - m_{ijk})^2}{m_{ijk}} \right]$$
- Kriteria penolakan  
Tolak  $H_0$  jika  $\chi_{hit}^2 > \chi_{tabel}^2$  atau  $p\_value \leq \alpha$
- Kesimpulan  
Jika  $\chi_{hit}^2 > \chi_{tabel}^2$  atau  $p\_value \leq \alpha$  maka  $H_0$  ditolak artinya model tidak sesuai dengan keadaan sebenarnya atau jika  $\chi_{hit}^2 \leq \chi_{tabel}^2$  atau  $p\_value > \alpha$  maka  $H_0$  diterima artinya model sesuai dengan keadaan sebenarnya

Atau dengan statistik uji *rasio likelihood* ( $G^2$ )

Dalam menentukan model terbaik harus memenuhi kriteria berikut:

- Memenuhi uji *goodness of fit*
- Mudah ditafsirkan atau diterjemahkan
- Model sesederhana mungkin

Dimulai dengan memilih nilai  $G^2$  atau ( $\chi^2$ ) yang relatif kecil. Namun jika terpilih beberapa model maka perlu dilakukan pemilihan model terbaik dengan partisi *chi-square*, guna mendapatkan model terbaik yang paling sederhana (Trastika 2006 dalam Lestyorini, 2010)

Menurut Agresti (2002), tujuan dari analisis residual adalah untuk mengukur sisa variabilitas data pengamatan yang tidak dapat dijelaskan baik dari masing-masing variabelnya maupun interaksi antar variabelnya, nilai residu yang terkecil diambil sebagai model log linier terbaik. Analisis residual dapat dicari dengan :

$$e_{ijk} = n_{ijk} - \hat{m}_{ijk}; \text{ untuk } i = 1 \dots I, j = 1 \dots J, k = 1 \dots K \quad (9)$$

### 3. Prosedur Model log Linier Pada Tabel Kontingensi Tak Sempurna

Tabel kontingensi empat dimensi merupakan tabel yang menggunakan empat variabel mempunyai konsep yang sama dengan tabel kontingensi dua dimensi dan tiga dimensi, hanya saja variabel yang digunakan lebih banyak yakni variabel  $W, X, Y$  dan  $Z$ . Tabel kontingensi dikatakan tabel kontingensi tak sempurna jika dan hanya jika tabel tersebut mengandung satu sel kosong atau lebih untuk data yang akan diamati (Destyarisani, 2013). Tabel kontingensi empat dimensi yang mempunyai  $(i \times j \times k \times l)$  sel, terdiri dari  $i$  baris,  $j$  kolom,  $k$  layer dan  $l$  lapis yang disebut sebagai tabel kontingensi  $i \times j \times k \times l$ . Tabel kontingensi  $i \times j \times k \times l$  dapat dilihat pada Tabel 3.1 berikut:

Tabel 4. Tabel kontingensi  $i \times j \times k \times l$

Variabel (W)	Variabel (X)	Variabel (Y)	Variabel 4 (Z)			Total
			Z <sub>1</sub>	...	Z <sub>l</sub>	
W <sub>1</sub>	X <sub>1</sub>	Y <sub>1</sub>	n <sub>1111</sub>	...	n <sub>111l</sub>	n <sub>111+</sub>
		⋮	⋮	...	⋮	⋮
		Y <sub>k</sub>	n <sub>11k1</sub>	...	n <sub>11kl</sub>	n <sub>11k+</sub>
	X <sub>j</sub>	Y <sub>1</sub>	n <sub>1j11</sub>	...	n <sub>1j1l</sub>	n <sub>1j1+</sub>
		⋮	⋮	...	⋮	⋮
		Y <sub>k</sub>	n <sub>1jk1</sub>	...	n <sub>1jkl</sub>	n <sub>1jk+</sub>
⋮	⋮	⋮	⋮	⋮	⋮	
W <sub>i</sub>	X <sub>1</sub>	Y <sub>1</sub>	n <sub>i111</sub>	...	n <sub>i11l</sub>	n <sub>i11+</sub>
		⋮	⋮	...	⋮	⋮
		Y <sub>k</sub>	n <sub>i1k1</sub>	...	n <sub>i1kl</sub>	n <sub>i1k+</sub>
	X <sub>j</sub>	Y <sub>1</sub>	n <sub>ij11</sub>	...	n <sub>ij1l</sub>	n <sub>ij1+</sub>
		⋮	⋮	...	⋮	⋮
		Y <sub>k</sub>	n <sub>ijk1</sub>	...	n <sub>ijkl</sub>	n <sub>ijk+</sub>
Total			n <sub>++++1</sub>	...	n <sub>++++l</sub>	n <sub>++++</sub>

Model *independent* untuk empat variabel adalah  $\log m_{ijkl} = \mu + \lambda_i^W + \lambda_j^X + \lambda_k^Y + \lambda_l^Z \quad (10)$

Untuk  $i = 1 \dots I, j = 1 \dots J, k = 1 \dots K, l = 1 \dots L$

Dimana :

$\mu$  : logaritma jumlah nilai harapannya atau rata-rata dari seluruh logaritma nilai harapannya.

$\lambda_i^W$  : pengaruh dari variabel 1 terhadap model.

$\lambda_j^X$  : pengaruh dari variabel 2 terhadap model.

$\lambda_k^Y$  : pengaruh dari variabel 3 terhadap model.

$\lambda_l^Z$  : pengaruh dari variabel 4 terhadap model.

dengan derajat bebasnya adalah

$$\begin{aligned} db &= db(\log m_{ijkl}) - (db(\mu) + db(\lambda_i^W) + db(\lambda_j^X) + db(\lambda_k^Y) + db(\lambda_l^Z)) \\ &= IJKL - ((1) + (I - 1) + (J - 1) + (K - 1) + (L - 1)) \\ &= IJK - I - J - K - L + 3 \end{aligned}$$

Jika terdapat interaksi pada keempat variabel, maka model menjadi

$$\begin{aligned} \log m_{ijkl} &= \mu + \lambda_i^W + \lambda_j^X + \lambda_k^Y + \lambda_l^Z + \lambda_{ij}^{WX} + \lambda_{ik}^{WY} \\ &+ \lambda_{il}^{WZ} + \lambda_{jk}^{XY} + \lambda_{jl}^{XZ} + \lambda_{kl}^{YZ} \\ &+ \lambda_{ijk}^{WXY} + \lambda_{ijl}^{WXZ} + \lambda_{jkl}^{XYZ} + \lambda_{ikl}^{WYZ} \\ &+ \lambda_{ijkl}^{WXYZ} \quad (11) \end{aligned}$$

untuk  $i = 1 \dots I, j = 1 \dots J, k = 1 \dots K, l = 1 \dots L$

Keterangan :

$m_{ijkl}$  : frekuensi nilai pada sel ke-  $ijkl$  dalam model

$\mu$  : pengaruh rata-rata

$\lambda_i^W$  : pengaruh dari variabel 1 terhadap model.

$\lambda_j^X$  : pengaruh dari variabel 2 terhadap model.

$\lambda_k^Y$  : pengaruh dari variabel 3 terhadap model.

$\lambda_l^Z$  : pengaruh dari variabel 4 terhadap model.

$\lambda_{ij}^{WX}$  : pengaruh interaksi dari variabel 1 dan variabel ke-2

$\lambda_{ik}^{WY}$  : pengaruh interaksi dari variabel 1 dan variabel ke-3

$\lambda_{il}^{WZ}$  : pengaruh interaksi dari variabel 1 dan variabel ke-4

$\lambda_{jk}^{XY}$  : pengaruh interaksi dari variabel 2 dan variabel ke-3

$\lambda_{jl}^{XZ}$  : pengaruh interaksi dari variabel 2 dan variabel ke-4

$\lambda_{kl}^{YZ}$  : pengaruh interaksi dari variabel 3 dan variabel ke-4

$\lambda_{ijk}^{WXY}$  : pengaruh interaksi dari variabel 1, variabel 2 dan variabel ke-3

$\lambda_{ijl}^{WXZ}$  : pengaruh interaksi dari variabel 1, variabel 2 dan variabel ke-4

$\lambda_{jkl}^{XYZ}$  : pengaruh interaksi dari variabel 2, variabel 3 dan variabel ke-4

$\lambda_{ikl}^{WYZ}$  : pengaruh interaksi dari variabel 1, variabel 3 dan variabel ke-4

$\lambda_{ijkl}^{WXYZ}$  : pengaruh interaksi dari variabel 1, variabel 2, variabel 3 dan variabel ke-4

Prosedur Analisis Model Log Linier Empat Dimensi

Menurut Hapsari (2011), statistik cukup minimal untuk model-model *log linier* merupakan koefisien dari masing-masing parameternya yang diperoleh dari pengumpulan dan penjumlahan batas marjinal dari masing-masing parameternya.

Tabel 5. Statistik Cukup Minimal untuk Model *Log Linier* Empat Dimensi

Simbol Model Log Linier	Statistik Cukup Minimal
(WXYZ)	$\{n_{ijkl}\}$
(WX, Y, Z)	$\{n_{ij++}, \{n_{++k+}, \{n_{+++l}\}$
(WY, X, Z)	$\{n_{i+k+}, \{n_{+j++}, \{n_{+++l}\}$
(WZ, X, Y)	$\{n_{i++l}, \{n_{+j++}, \{n_{++k+}\}$
(XY, W, Z)	$\{n_{+jk+}, \{n_{i+++}, \{n_{+++l}\}$
(XZ, W, Y)	$\{n_{+j+l}, \{n_{i+++}, \{n_{++k+}\}$
(YZ, W, X)	$\{n_{++kl}, \{n_{i+++}, \{n_{+j++}\}$
(WX, YZ)	$\{n_{ij++}, \{n_{++kl}\}$
(WX, XY, Z)	$\{n_{ij++}, \{n_{+jk+}, \{n_{+++l}\}$
(WX, WY, WZ)	$\{n_{ij++}, \{n_{i+k+}, \{n_{i+++}\}$
(WX, XY, YZ)	$\{n_{ij++}, \{n_{+jk+}, \{n_{++kl}\}$
(WX, WY, XY, Z)	$\{n_{ij++}, \{n_{i+k+}, \{n_{+jk+}, \{n_{+++l}\}$
(WX, WY, XY, YZ)	$\{n_{ij++}, \{n_{i+k+}, \{n_{+jk+}, \{n_{++kl}\}$
(WX, WY, XY, XZ)	$\{n_{ij++}, \{n_{i+k+}, \{n_{+jk+}, \{n_{+j+l}, \{n_{++kl}\}$
(WXYZ, Z)	$\{n_{ijk+}, \{n_{+++l}\}$
(WXY, YZ)	$\{n_{ijk+}, \{n_{++kl}\}$
(WXY, XZ, YZ)	$\{n_{ijk+}, \{n_{+j+l}, \{n_{++kl}\}$
(WXY, WY, XZ)	$\{n_{ijk+}, \{n_{i+k+}, \{n_{+j+l}\}$
(WXY, WY, XY, YZ)	$\{n_{ijk+}, \{n_{i+k+}, \{n_{+jk+}, \{n_{++kl}\}$
(WXY, WZ, XZ)	$\{n_{ijk+}, \{n_{i++l}, \{n_{+j+l}, \{n_{++kl}\}$
(WXY, XYZ)	$\{n_{ijk+}, \{n_{+jkl}\}$
(WXY, WXZ, YZ)	$\{n_{ijk+}, \{n_{ij+l}, \{n_{++kl}\}$
(WXZ, XYZ, WY)	$\{n_{ij+l}, \{n_{+jkl}, \{n_{i+++}\}$
(WXY, WXZ, XY)	$\{n_{ijk+}, \{n_{ij+l}, \{n_{+jkl}\}$
(WXZ, WYZ, XY)	$\{n_{i++l}, \{n_{i++kl}, \{n_{+jkl}\}$
(WXY, WXZ, XY, YZ)	$\{n_{ijk+}, \{n_{ij+l}, \{n_{+jkl}, \{n_{i++kl}\}$
(W, X, Y, Z)	$\{n_{i+++}, \{n_{+j++}, \{n_{++k+}, \{n_{+++l}\}$

Uji *goodness of fit* pada tabel kontingensi dilakukan untuk melihat apakah ada perbedaan antara frekuensi observasi dengan frekuensi yang diharapkan, pada tabel kontingensi tak sempurna berdimensi empat dalam penelitian ini, uji *goodness of fit* hanya dilihat dari nilai *pearson chi-square* saja, karena pada tabel kontingensi tak sempurna mempunyai beberapa sel kosong atau bernilai 0, sehingga jika dilakukan dengan melihat nilai statistik *rasio likelihood* maka nilai dari

statistik *rasio likelihood* ( $G^2$ ) tidak akan terdefinisi karena nilai  $\ln 0$  adalah tak hingga, adapun langkah-langkah pengujiannya sebagai berikut (Agesti, 2007):

- Hipotesis  
 $H_0$  : Model sesuai dengan keadaan sebenarnya  
 $H_1$  : Model tidak sesuai dengan keadaan sebenarnya
- Taraf signifikansi  
 $\alpha = 0.05$
- Statistik Uji :  

$$\chi^2 = \sum_{ijkl} \left[ \frac{(n_{ijkl} - m_{ijkl})^2}{m_{ijkl}} \right]$$
- Kriteria penolakan  
Tolak  $H_0$  jika  $\chi_{hit}^2 > \chi_{tabel}^2$  atau  $p\_value \leq \alpha$
- Kesimpulan  
Jika  $\chi_{hit}^2 > \chi_{tabel}^2$  atau  $p\_value \leq \alpha$  maka  $H_0$  ditolak artinya model tidak sesuai dengan keadaan sebenarnya atau jika  $\chi_{hit}^2 \leq \chi_{tabel}^2$  atau  $p\_value > \alpha$  maka  $H_0$  diterima artinya model sesuai dengan keadaan sebenarnya

Dalam menentukan model terbaik harus memenuhi kriteria berikut (Hapsari, 2011):

- Memenuhi uji *goodness of fit*
- Mudah ditafsirkan atau diterjemahkan
- Model sesederhana mungkin

Pada tabel kontingensi tak sempurna berdimensi empat ini, prosedur pemilihan model terbaik dimulai dengan memilih nilai  $\chi^2$  yang relatif kecil. Selain itu harus dilakukan Partisi *chi-square* untuk mendapatkan model terbaik yang paling sederhana (Trastika 2006 dalam Lestyorini, 2010).

Analisis residual ini bertujuan untuk mengukur sisa variabilitas data pengamatan yang tidak dapat dijelaskan baik dari masing-masing variabelnya maupun interaksi antar variabelnya, nilai residu yang terkecil diambil sebagai model log linier terbaik. Analisis residual dapat dicari dengan (Agesti, 2002):

$$e_{ijkl} = n_{ijkl} - \hat{m}_{ijkl} \text{ untuk } i = 1 \dots I, j = 1 \dots J, k = 1 \dots K, l = 1 \dots L \quad (11)$$

#### 4. Analisis Data Pada Tabel Kontingensi Tak Sempurna Berdimensi Empat

Data yang digunakan pada penelitian ini berupa data sekunder. Data penelitian ini yaitu data jumlah pegawai negeri sipil Provinsi Jawa Barat tahun 2012 yang diambil dari kantor Badan Pusat Statistik (BPS) kota Bengkulu (Lampiran 2). Deskripsi dari data penelitian tersebut adalah sebagai berikut :

- Variabel 1 adalah unit kerja ( $W$ ) untuk variabel kolom, dengan  $i = 1, \dots, 28$  merupakan jumlah dari kolom
- Variabel 2 adalah pendidikan formal ( $X$ ) untuk variabel baris, dengan  $j = 1, 2$  merupakan jumlah dari baris

- Variabel 3 adalah jenis kelamin ( $Y$ ) untuk variabel layer, dengan  $k = 1, 2$  merupakan jumlah dari layer
- Variabel 4 adalah golongan ( $Z$ ) untuk variabel lapis, dengan  $k = 1, 2, 3, 4$  merupakan jumlah dari lapis

Prosedur analisis pada data Jumlah Pegawai Negeri Sipil Provinsi Jawa Barat Tahun 2012 berdasarkan unit kerja/institusi ( $W$ ), tingkat pendidikan formal ( $X$ ), jenis kelamin ( $Y$ ) dan golongan ( $Z$ ), yang disajikan pada tabel kontingensi tak sempurna berdimensi empat (Lampiran 3) adalah sebagai berikut:

Mencari statistik cukup minimal untuk data Jumlah Pegawai Negeri Sipil Provinsi Jawa Barat Tahun 2012 berdasarkan unit kerja/institusi ( $W$ ), tingkat pendidikan formal ( $X$ ), jenis kelamin ( $Y$ ) dan golongan ( $Z$ ).

Data yang digunakan dalam penelitian ini disajikan dalam tabel kontingensi, dimana beberapa sel pada tabel tersebut bernilai nol, maka tabel tersebut disebut sebagai tabel kontingensi tak sempurna, untuk uji *goodness of fit* yang dilakukan hanya melihat nilai dari statistik *pearson chi-square* saja, karena nilai dari statistik *rasio likelihood* ( $G^2$ ) tidak akan terdefinisi. Berdasarkan hasil analisis yang diperoleh dari data penelitian, pada model ( $W, X, Y, Z$ ) yang berarti bahwa ke empat faktor yang mempengaruhi data saling bebas atau tidak ada interaksi antara ke empat faktor yang mempengaruhi data, diperoleh Tabel 4.2 sebagai berikut:

Tabel 6. Tabel Analisis Model Saling Bebas

Model	Model Log Linier	DB	$(\chi^2)$	$(\chi^2)$ tabel	Keterangan
1	( $W, X, Y, Z$ )	415	14315.98	463.497	$H_0$ Ditolak

Berdasarkan tabel di atas, model ( $W, X, Y, Z$ ) dengan DB 415, dimana  $(\chi^2_{hit}) > (\chi^2_{tabel})$  atau  $14316 > 463.497$  maka dapat disimpulkan bahwa  $H_0$  ditolak artinya model ( $W, X, Y, Z$ ) tidak sesuai dengan keadaan sebenarnya.

Selanjutnya dilakukan analisis untuk dua faktor yang saling berinteraksi dan dua lainnya saling bebas, dimana untuk model dengan dua faktor yang saling berinteraksi dan dua lainnya saling bebas akan menghasilkan sebanyak 6 model, model tersebut dapat dilihat pada Tabel 7 berikut:

Tabel 7. Tabel Analisis Model Untuk 2 Faktor Berinteraksi dan 2 Faktor Lainnya Saling Bebas

Mode l	Model Log Linier	DB	$(\chi^2)$	$(\chi^2)$ tabel	Keteranga n
2	( $WX, Y, Z$ )	388	9259.2737	434.9288	$H_0$ Ditolak
3	( $WY, X, Z$ )	388	10716.04	434.9288	$H_0$ Ditolak
4	( $WZ, X, Y$ )	390	9435.23	437.0468	$H_0$ Ditolak
5	( $XY, W, Z$ )	414	13630.30	462.4401	$H_0$ Ditolak
6	( $XZ, W, Y$ )	412	14052.56	460.3255	$H_0$ Ditolak
7	( $YZ, W, X$ )	412	13460.638	460.3255	$H_0$ Ditolak

Untuk model dengan dua faktor yang saling berinteraksi dan dua lainnya saling bebas, dengan melihat keadaan yang sebenarnya bahwa diantara faktor-faktor yang mempengaruhi data yaitu unit

kerja, tingkat pendidikan formal, jenis kelamin dan golongan, dapat dikatakan bahwa tingkat pendidikan formal akan mempengaruhi golongan, dengan melihat model 5 yaitu ( $XZ, W, Y$ ) dimana nilai DBnya adalah 414 dan nilai  $(\chi^2_{hit}) > (\chi^2_{tabel})$  atau  $13630.30 > 462.4401$  maka dapat disimpulkan bahwa  $H_0$  ditolak artinya model ( $XZ, W, Y$ ) tidak sesuai dengan keadaan sebenarnya.

Selanjutnya untuk model dua faktor saling berinteraksi dan dua lainnya saling bebas, pada model 2 yaitu ( $WX, Y, Z$ ) dengan menganggap bahwa faktor 1 mempengaruhi faktor ke 2, dengan DB sebesar 388 dan  $(\chi^2_{hit}) > (\chi^2_{tabel})$  atau  $9259.2737 > 434.9288$  maka dapat disimpulkan bahwa  $H_0$  ditolak artinya model ( $XZ, W, Y$ ) tidak sesuai dengan keadaan sebenarnya. Pada model 3 yaitu ( $WX, Y, Z$ ) dengan menganggap bahwa faktor 1 mempengaruhi faktor ke 3, dengan DB sebesar 388 dan  $(\chi^2_{hit}) > (\chi^2_{tabel})$  atau  $10716.04 > 434.9288$  maka dapat disimpulkan bahwa  $H_0$  ditolak artinya model ( $WX, Y, Z$ ) tidak sesuai dengan keadaan sebenarnya. Pada model 4 yaitu ( $WZ, X, Y$ ) dengan menganggap bahwa faktor 1 mempengaruhi faktor ke 4, dengan DB sebesar 390 dan  $(\chi^2_{hit}) > (\chi^2_{tabel})$  atau  $9435.23 > 437.0468$  maka dapat disimpulkan bahwa  $H_0$  ditolak artinya model ( $WZ, X, Y$ ) tidak sesuai dengan keadaan sebenarnya. Pada model 6 yaitu ( $XZ, W, Y$ ) dengan menganggap bahwa faktor 2 mempengaruhi faktor ke 4, dengan DB sebesar 412 dan  $(\chi^2_{hit}) > (\chi^2_{tabel})$  atau  $14052.56 > 460.3255$  maka dapat disimpulkan bahwa  $H_0$  ditolak artinya model ( $XZ, W, Y$ ) tidak sesuai dengan keadaan sebenarnya. Pada model 7 yaitu ( $YZ, W, X$ ) dengan menganggap bahwa faktor 2 mempengaruhi faktor ke 4, dengan DB sebesar 412 dan  $(\chi^2_{hit}) > (\chi^2_{tabel})$  atau  $13460.638 > 460.3255$  maka dapat disimpulkan bahwa  $H_0$  ditolak artinya model ( $YZ, W, X$ ) tidak sesuai dengan keadaan sebenarnya.

Untuk analisis selanjutnya karena untuk tabel kontingensi berdimensi empat jika ingin melihat semua kemungkinan model yang ada, maka akan menghasilkan kemungkinan model yang sangat banyak, sehingga untuk analisis selanjutnya akan dibahas beberapa kemungkinan model saja yang saling berinteraksi. Berdasarkan analisis data yang dilakukan diperoleh hasil pada tabel berikut:

Tabel 8. Tabel Analisis Model Lainnya

Model	Model Log Linier	Db	$(\chi^2)$	$(\chi^2)$ tabel	Keterangan
8	( $WX, YZ$ )	385	8295.1386	431.75	$H_0$ Ditolak
9	( $WX, XY, Z$ )	387	24513.18	228.58	$H_0$ Ditolak
10	( $WX, WY, WZ$ )	384	29709165.24	430.69	$H_0$ Ditolak
11	( $WX, XY, YZ$ )	280	69849.46	320.03	$H_0$ Ditolak
12	( $WX, WY, XY$ )	361	396003560.9	406.31	$H_0$ Ditolak
13	( $WX, WY, XY$ )	357	5332362.4	402.059	$H_0$ Ditolak
14	( $WX, WY, XY$ )	354	2189988.65	398.87	$H_0$ Ditolak
15	( $WX, WY, XY$ )	273	5.96E-08	313	$H_0$ Diterima
16	( $WXY, Z$ )	333	5869.4	376.5	$H_0$ Ditolak
17	( $WXY, YZ$ )	330	16465.8	373.36	$H_0$ Ditolak

18	(WXY, XZ, YZ)	230	2377272.20	266.4	H <sub>0</sub> Ditolak
19	(WXY, WY, X)	276	707387.9	315.75	H <sub>0</sub> Ditolak
20	(WXY, WY, X)	228	992287.8	264.2	H <sub>0</sub> Ditolak
21	(WXY, WZ, X)	301	630491607.41	342.46	H <sub>0</sub> Ditolak
22	(WXY, XYZ)	324	4784.7	366.98	H <sub>0</sub> Ditolak
23	(WXY, WXZ)	112	11043657.03	137.7	H <sub>0</sub> Ditolak
24	(WXZ, XYZ, Y)	208	9523360.9	242.65	H <sub>0</sub> Ditolak
25	(WXY, WXZ)	129	2292580.1	156.51	H <sub>0</sub> Ditolak
26	(WXZ, WYZ)	108	127442504.6	133.26	H <sub>0</sub> Ditolak
27	(WXY, WXZ)	81	47951561704.2	103.01	H <sub>0</sub> Ditolak
28	(WXYZ)	0	0	0	H <sub>0</sub> Diterima

Berdasarkan Tabel 8 di atas dapat dilihat nilai statistik *Pearson Chi-Square* untuk model 7 sampai model 27 kecuali model 15, karena nilai dari  $(\chi^2_{hit}) > (\chi^2_{tabel})$  untuk model 7 sampai model 27 kecuali model 15, maka memberikan keputusan bahwa  $H_0$  ditolak, yang berarti bahwa model 7 sampai model 27 kecuali model 15 tidak sesuai dengan keadaan sebenarnya, sedangkan untuk nilai statistik *Pearson Chi-Square* pada model 15 dengan DB 273 dan nilai  $(\chi^2_{hit}) \leq (\chi^2_{tabel})$  atau  $5.96E-08 \leq 313$  maka memberikan keputusan bahwa  $H_0$  diterima, yang berarti bahwa model 15 sesuai dengan keadaan sebenarnya, selanjutnya pada model 28 dengan DB 0 dan nilai dari  $(\chi^2_{hit}) \leq (\chi^2_{tabel})$  atau  $0 \leq 0$ , maka memberikan keputusan bahwa  $H_0$  diterima, yang berarti bahwa model 28 sesuai dengan keadaan sebenarnya.

Tahap selanjutnya adalah analisis pemilihan model terbaik dari data Jumlah Pegawai Negeri Sipil Provinsi Jawa Barat Tahun 2012 yang merupakan analisis pemilihan model terbaik pada tabel kontingensi tak sempurna berdimensi empat, dimana pemilihan model terbaik menggunakan partisi *chi-square*, dimana pada tahap sebelumnya telah dilakukan uji *Goodness of Fit* dengan melihat nilai dari statistik *Pearson Chi-Square* dari masing-masing model dengan ketentuan  $\chi^2_{hit} \leq \chi^2_{tabel}$  maka uji *Goodness of Fi* terpenuhi, adapun model yang memenuhi uji *Goodness of Fit* dapat dilihat pada tabel berikut:

Tabel 9. Tabel Pemilihan Model

Model	Model Log Linier	Db	( $\chi^2$ )	Keterangan
15	(WX, WY, XY, XZ, YZ, WZ)	273	5.96E-08	H <sub>0</sub> Diterima
28	(WXYZ)	0	0	H <sub>0</sub> Diterima

Dalam pembahasan selanjutnya akan dibandingkan model yang terpilih dari tahap sebelumnya yaitu tahap pemilihan model terbaik, dengan melihat selisih nilai statistik *Pearson Chi-Square* dan selisih DB dari model terpilih, tabel dari partisi *Chi-Square* dapat dilihat sebagai berikut:

Tabel 10. Tabel Partisi *Chi - Square*

Model	Model Log Linier	Db	selisih	( $\chi^2$ )	selisih	Keterangan
15	(WX, WY, XY, XZ, YZ, WZ)	273	-	5.96E-08	-	Model 15 lebih baik dari model 28

28	(WXYZ)	0	-	0	-	-
----	--------	---	---	---	---	---

Berdasarkan tabel di atas dapat disimpulkan bahwa model yang terbaik adalah model 15, dengan nilai statistik *Pearson Chi-Square* ( $\chi^2$ ) sebesar 5.96E-08, sehingga model 15 adalah model yang dipilih, dimana modelnya adalah (WX, WY, XY, XZ, YZ, WZ). Pada model 15 diketahui bahwa model tersebut merupakan model dua arah yang berarti bahwa adanya hubungan antara dua faktor yang mempengaruhi data jumlah pegawai negeri sipil yakni, faktor unit kerja berinteraksi dengan pendidikan, unit kerja berinteraksi dengan jenis kelamin, unit kerja berinteraksi dengan golongan, pendidikan berinteraksi dengan jenis kelamin, pendidikan berinteraksi dengan golongan dan jenis kelamin berinteraksi dengan golongan. Dalam persamaan model *log linier* dapat ditulis sebagai berikut:

$$\log m_{ijk} = \mu + \lambda_i^W + \lambda_j^X + \lambda_k^Y + \lambda_l^Z + \lambda_{ij}^{WX} + \lambda_{ik}^{WY} + \lambda_{jk}^{XY} + \lambda_{jl}^{XZ} + \lambda_{kl}^{YZ} + \lambda_{il}^{WZ}$$

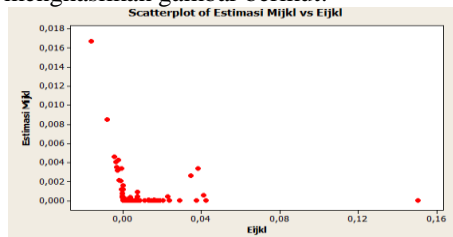
Keterangan :

- $m_{ijkl}$  : frekuensi nilai pada sel ke- *ijkl* dalam model
- $\mu$  : pengaruh rata-rata
- $\lambda_i^W$  : pengaruh dari variabel 1 terhadap model.
- $\lambda_j^X$  : pengaruh dari variabel 2 terhadap model.
- $\lambda_k^Y$  : pengaruh dari variabel 3 terhadap model.
- $\lambda_l^Z$  : pengaruh dari variabel 4 terhadap model.
- $\lambda_{ij}^{WX}$  : pengaruh interaksi dari variabel unit kerja dan variabel pendidikan formal
- $\lambda_{ik}^{WY}$  : pengaruh interaksi dari variabel unit kerja dan variabel jenis kelamin
- $\lambda_{il}^{XY}$  : pengaruh interaksi dari variabel unit kerja dan variabel golongan
- $\lambda_{jk}^{XZ}$  : pengaruh interaksi dari variabel pendidikan formal dan variabel jenis kelamin
- $\lambda_{jl}^{YZ}$  : pengaruh interaksi dari variabel pendidikan formal dan variabel golongan
- $\lambda_{kl}^{WZ}$  : pengaruh interaksi dari variabel jenis kelamin dan variabel golongan

Dalam penelitian ini model terbaik adalah model dengan simbol (WX, WY, XY, XZ, YZ, WZ), sehingga dilakukan analisis residual pada model tersebut yang bertujuan untuk mengukur sisa variabilitas data pengamatan, dimana residual merupakan nilai dari frekuensi pengamatan dikurangi dengan frekuensi harapan atau dilambangkan dengan  $e_{ijkl} = n_{ijkl} - \hat{m}_{ijkl}$  untuk  $i = 1 \dots I, j = 1 \dots J, k = 1 \dots K, l = 1 \dots L$ .

Nilai residual positif mempunyai arti bahwa frekuensi pengamatan lebih besar dari frekuensi harapan, sebaliknya jika frekuensi harapan lebih

besar dari frekuensi pengamatan maka nilai residualnya negatif. Jika nilai residual diplotkan dengan nilai estimasi frekuensi harapan dengan menggunakan program minitab maka akan menghasilkan gambar berikut:



**Gambar 1. Scatterplot Nilai Residual Berdasarkan Estimasi Nilai Harapan**

Berdasarkan Gambar 1. di atas menunjukkan bahwa nilai residualnya relatif kecil (mendekati nilai nol), sehingga model dengan simbol  $(WX, WY, XY, XZ, YZ, WZ)$  adalah model terbaik untuk mewakili data dengan kasus jumlah pegawai negeri sipil Provinsi Jawa Barat Tahun 2012.

### 5. Kesimpulan dan Saran

Berdasarkan hasil pembahasan mengenai model *log linier* pada tabel kontingensi tak sempurna empat dimensi beserta penerapannya dalam kasus Jumlah Pegawai Negeri Sipil Provinsi Jawa Barat tahun 2012, dapat diambil kesimpulan sebagai berikut:

1. Model *log linier* empat dimensi merupakan perluasan dari model *log linier* tiga dimensi. Model *log linier* pada tabel kontingensi tak sempurna adalah model *log linier* yang dilakukan pada tabel yang mempunyai sebuah sel kosong atau lebih untuk populasi yang ditinjau dengan dimensi empat. Bentuk umum model *log linier* pada tabel kontingensi tak sempurna empat dimensi yaitu:
$$\log m_{ijkl} = \mu + \lambda_i^W + \lambda_j^X + \lambda_k^Y + \lambda_l^Z + \lambda_{ij}^{WX} + \lambda_{ik}^{WY} + \lambda_{il}^{XY} + \lambda_{jk}^{XZ} + \lambda_{jl}^{YZ} + \lambda_{kl}^{WZ} + \lambda_{ijk}^{WXY} + \lambda_{ijl}^{WXZ} + \lambda_{jkl}^{XYZ} + \lambda_{ijkl}^{WXYZ}$$
2. Penerapan model *log linier* pada tabel kontingensi tak sempurna berdimensi empat pada data Jumlah Pegawai Negeri Sipil Provinsi Jawa Barat Tahun 2012. Berdasarkan hasil analisis data, diperoleh 28 model yang mungkin terjadi untuk model *log linier* pada tabel kontingensi tak sempurna berdimensi empat. Pada analisis data ada 2 model yang memenuhi uji *Goodness of Fit*, kemudian dilakukan analisis lebih lanjut dengan pemilihan model

terbaik menggunakan partisi *Chi-Square*, sehingga menghasilkan satu model terbaik dengan nilai  $(\chi^2)$  terkecil. Oleh sebab itu model *log linier* terbaik pada kasus data Jumlah Pegawai Negeri Sipil Provinsi Jawa Barat Tahun 2012 adalah  $\log m_{ijk} = \mu + \lambda_i^W + \lambda_j^X + \lambda_k^Y + \lambda_l^Z + \lambda_{ij}^{WX} + \lambda_{ik}^{WY} + \lambda_{jk}^{XY} + \lambda_{jl}^{XZ} + \lambda_{kl}^{XZ} + \lambda_{il}^{WZ}$  yang disimbolkan dengan  $(WX, WY, XY, XZ, YZ, WZ)$ , dimana model ini juga memiliki nilai residual yang relatif kecil (mendekati nilai nol). Kesimpulan dari model terbaik tersebut bahwa variabel unit kerja ( $W$ ) berinteraksi dengan ketiga variabel yang lain yaitu pendidikan formal ( $X$ ), jenis kelamin ( $Y$ ), golongan ( $Z$ ), pendidikan formal ( $X$ ) berinteraksi dengan kedua variabel lain yaitu, jenis kelamin ( $Y$ ), golongan ( $Z$ ) dan variabel jenis kelamin ( $Y$ ) berinteraksi dengan variabel golongan ( $Z$ ). Arti dari model terbaik tersebut adalah perbedaan unit kerja mempengaruhi ketiga variabel yang lain yaitu, pendidikan formal, jenis kelamin dan golongan, perbedaan pendidikan formal mempengaruhi kedua variabel yang lain yaitu jenis kelamin, dan golongan, serta variabel jenis kelamin mempengaruhi golongan

Penulis hanya melakukan analisis model *log linier* pada tabel kontingensi tak sempurna berdimensi empat dan menerapkannya pada kasus data Jumlah Pegawai Negeri Sipil Provinsi Jawa Barat Tahun 2012 dengan melihat nilai dari statistik *Pearson Chi-Square* saja, sehingga pembaca dapat melakukan analisis untuk model *log linier* pada tabel kontingensi tak sempurna dengan melihat nilai statistik *rasio likelihood* ( $G^2$ ) dengan memperhatikan data dan melakukan penggabungan antara baris, kolom, layer atau lapis pada data yang mengandung nilai nol. Penggabungan antara baris, kolom, layer atau lapis hanya dapat dilakukan jika jumlah dari variabel kategorik baris ( $i$ ), kolom ( $j$ ), layer ( $k$ ) dan lapis ( $l$ ) lebih dari dua kategori.

### 6. Daftar Pustaka

- Agresti, A. 2002. *Categorical Data Analysis*. John Willey and Sons. USA.
- Agresti, A. 2007. *An Introduction to Categorical Data Analysis*. John Willey and Sons. USA.
- Anonim. 2012. Badan Pusat Statistik Provinsi Bengkulu. Jawa Barat Dalam Angka Tahun 2007-2012. Bengkulu
- Brzezinska, Justyna. 2012. *Hierarchical Log Linier Models For Contingency Tables*. Universitas Lodzensis.
- Destyarisani, Tiara. *Model Log Linier Jumlah Narapidana Kota Bengkulu Tahun 2011*. Fakultas MIPA UNIB. Bengkulu. Tidak Dipublikasikan.
- Fienberg, S.E. 2007. *The Analysis Cross-Classified Categorical Data, Second Edition*. Springer Science + Business Media. USA.
- Gottard, A., and Marchetti, G.M. 2011. *Quasi-Symmetric Graphical Log Linier Models*.

- Scandinavian Journal of Statistic*. Vol. 38. 447-465.
- Hapsari, G.S. 2011. *Model Log Linier Untuk Tabel Kontingensi Tak Sempurna Berdimensi Tiga (Studi Kasus Jumlah Penduduk Kabupaten Sleman Tahun 2008 Menurut Umur, Pendidikan dan Jenis Kelamin)*. Fakultas MIPA UNY. Yogyakarta.
- Indratno, I., dan Irwinsyah, R. 1998. Aplikasi Analisis Tabulasi Silang (Crosstab) Dalam Perencanaan Wilayah dan Kota. *Jurnal PWK*. Vol. 9, No. 2, 48-59.
- Kamil, I., Susilawati, M., and Kencana, I. P.E. Model Log Linier Faktor-Faktor yang Mempengaruhi Hipertensi. *E-jurnal Matematika*. Vol. 1, No. 1, 84-88
- Klimova, A., Rudas, T., and Dobra, A. 2011. *Relational Models For Contingency Tables*. University of Washington.
- Kusumawardani, I.S., Gumila, I., dan Rostini, L. 2012. Analisis Surplus Konsumen dan Surplus Produsen Ikan Segar di Kota Bandung. *Jurnal Perikanan dan Kelautan*. Vol. 3, No. 4, 141-150.
- Lestiyorini, M. 2010. *Model Log Linear Multivariat Empat Dimensi (Studi Kasus : Akses Internet Mahasiswa Jurusan Pendidikan Matematika di Universitas Negeri Yogyakarta)*. Fakultas MIPA UNY. Yogyakarta.
- Nopitasari, R. 2012. *Analisis Data Kategorik Dengan Pengkelasan Tertata (Ordered Categories)*. Fakultas MIPA UNIB. Bengkulu. Tidak Dipublikasikan.
- Olmus, H., and Erbras, S. 2012. *Analysis of Traffic Accidents Caused by Drivers by Using Log-Linier Models*. *Promet-Traffic & Transportation*. Vol. 24, No. 6. 495-504.
- Rahmana, A.M., dan Rahayu, S.P. (2013). Analisis Pola Hubungan Kerugian Negara Akibat Korupsi dengan Demografi Koruptor di Jawa Timur. *Jurnal Sains dan Semi Pomits*. Vol. 2, No. 2, 305-310.
- Wulandari, S.P., Salamah, M., dan Susilaningrum, D. 2009. *Diktat Pengajaran Analisis Data Kualitatif*. Fakultas Mipa Institut Teknologi Sepuluh Nopember. Surabaya.

# Model Regresi Multivariat Terbaik Untuk Mengetahui Faktor-Faktor Yang Mempengaruhi Derajat Kesehatan di Indonesia

Riyan Hidayat

Jurusan Matematika Fakultas MIPA Universitas Bengkulu

Email : Riyan\_Studio07@yahoo.com

## ABSTRAK

Model regresi multivariat merupakan model regresi linier dengan lebih dari satu variabel *dependent Y* yang saling berkorelasi dan satu atau lebih variabel *independent X*. Pada analisis model regresi multivariat pemilihan model terbaik merupakan suatu hal penting. Hal ini dikarenakan pemilihan model terbaik pada model regresi multivariat tergantung banyaknya variabel *independent* yang terlibat dalam model. Tujuan penelitian ini yaitu untuk mengetahui faktor-faktor apa saja yang mempengaruhi derajat kesehatan di Indonesia dan dapat menentukan model regresi multivariat terbaik berdasarkan faktor-faktor yang mempengaruhi derajat kesehatan di Indonesia. Berdasarkan hasil pemilihan model menggunakan metode  $AIC_C$  diperoleh faktor-faktor yang mempengaruhi derajat kesehatan di Indonesia yaitu persentase kelahiran yang dilakukan oleh tenaga medis, persentase imunisasi lengkap dan persentase rumah tangga menggunakan fasilitas jamban sendiri.

Kata Kunci : Model Regresi Multivariat, Derajat Kesehatan,  $AIC_C$

## PENDAHULUAN

Pada era globalisasi saat ini, kesejahteraan adalah salah satu aspek yang penting dalam kehidupan bermasyarakat yang dapat dilihat dari derajat kesehatan penduduknya. Kesehatan merupakan investasi untuk mendukung pembangunan ekonomi serta memiliki peran penting dalam upaya penanggulangan kemiskinan.

Kondisi pembangunan kesehatan secara umum dapat dilihat dari status kesehatan dan gizi masyarakat, yaitu angka kematian bayi, prevalensi gizi buruk balita dan angka harapan hidup.

Indikator kesehatan di Indonesia pada setiap tahun memang menunjukkan perbaikan. Namun, laju perbaikan itu masih terbilang lambat. Situasi kesehatan di Indonesia masih tertinggal jauh dari Singapura, Malaysia, Thailand, dan Filipina.

Berdasarkan uraian diatas, faktor-faktor yang mempengaruhi derajat kesehatan di Indonesia sangat perlu diketahui sebagai bahan evaluasi pembangunan kesehatan agar Indonesia tidak tertinggal jauh dari negara-negara ASEAN lain dalam hal pembangunan. Metode statistika yang dapat digunakan untuk mengetahui faktor-faktor yang mempengaruhi derajat kesehatan di Indonesia adalah dengan menggunakan model regresi multivariat.

Salah satu metode dalam pemilihan model terbaik yang sangat terkemuka adalah  $AIC$

(*Akaike Information Criteria*). Walaupun  $AIC$  merupakan metode yang sudah terbukti secara luas, tetapi hal ini masih mempunyai kekurangan. Pada pemodelan regresi, Hurvich dan Tsai (1989) menunjukkan bahwa ketika sampel berukuran kecil, atau ketika jumlah parameter yang pas menjadi sedikit lebih besar dari ukuran kecil, metode  $AIC$  menjadi metode yang kurang pas untuk pemilihan model. Selanjutnya mereka merumuskan metode baru sebagai versi perbaikan dari metode  $AIC$  yang disebut sebagai  $AIC_C$ . Pada pemilihan model menggunakan metode  $AIC_C$  cenderung lebih baik dan lebih mengurangi bias dari pada pemilihan model menggunakan  $AIC$ .

Pemakaian model regresi multivariat ini mengacu pada penelitian yang pernah dilakukan oleh (Riskiyanti dan Wulandari, 2010) dengan judul *Analisis Regresi Multivariat Berdasarkan Faktor-Faktor yang Mempengaruhi Derajat Kesehatan di Provinsi Jawa Timur*.

Berdasarkan uraian dan juga acuan dari penelitian yang telah dilakukan sebelumnya, peneliti bermaksud untuk melanjutkan penelitian tersebut dengan judul "**Model Regresi Multivariat Terbaik Untuk Mengetahui Faktor-Faktor yang Mempengaruhi Derajat Kesehatan di Indonesia Tahun 2013**" dengan tetap menggunakan faktor-faktor yang berpengaruh pada penelitian sebelumnya dan menambahkan faktor-faktor lain yang tidak dimasukkan pada penelitian sebelumnya serta pada penelitian ini peneliti akan menggunakan metode  $AIC_C$  sebagai

metode yang digunakan pada pemilihan model terbaiknya.

Serta tujuan yang ingin dicapai dalam penelitian ini adalah untuk mengetahui faktor-faktor apa saja yang mempengaruhi derajat kesehatan di Indonesia dan dapat membuat model regresi multivariat terbaik berdasarkan faktor-faktor yang mempengaruhi derajat kesehatan di Indonesia.

## TINJAUAN PUSTAKA

### Definisi Derajat Kesehatan

Kesehatan adalah suatu keadaan fisik, mental, dan sosial yang terbebas dari suatu penyakit sehingga seseorang dapat melakukan aktifitas secara optimal.

Indikator utama yang digunakan untuk melihat derajat kesehatan penduduk adalah angka kematian bayi, angka harapan hidup, dan status gizi buruk balita (Anonim, 2015a).

#### Angka Kematian Bayi

Angka kematian bayi adalah angka yang menunjukkan banyaknya kematian bayi usia 0 tahun dari setiap 1000 kelahiran hidup pada tahun tertentu atau dapat dikatakan juga sebagai probabilitas bayi meninggal sebelum mencapai usia satu tahun (dinyatakan dengan per seribu kelahiran hidup) (Anonim, 2014b).

#### Angka Harapan Hidup

Angka harapan hidup pada suatu umur  $X$  adalah rata-rata tahun hidup yang dijalani oleh seseorang yang telah berhasil mencapai umur  $X$ , pada suatu tahun tertentu, dalam situasi mortalitas yang berlaku di lingkungan masyarakat (Anonim, 2014b).

#### Status Gizi Buruk Balita

Status gizi balita dapat diukur berdasarkan umur, berat badan (BB) dan tinggi badan (TB). Variabel umur, BB dan TB ini disajikan dalam bentuk tiga indikator *atropometri*, yaitu : berat badan menurut umur (BB/U), tinggi badan menurut umur (TB/U) dan berat badan menurut tinggi badan (BB/TB) (Anonim, 2013).

### Faktor-faktor yang Mempengaruhi Derajat Kesehatan

Menurut Blum (1984) ada 4 faktor yang mempengaruhi derajat kesehatan masyarakat atau perorangan. Faktor-faktor tersebut antara lain sebagai berikut :

1. Lingkungan
2. Perilaku
3. Pelayanan Kesehatan
4. Kependudukan/keturunan

### Model Regresi Multivariat

Model regresi linier multivariat adalah model regresi linier dengan lebih dari satu variabel *dependent*  $Y$  yang saling berkorelasi dan satu atau lebih variabel *independent*  $X$  (Johnson dan Wichern, 2002).

Dengan model sebagai berikut :

$$Y_{(n \times p)} = X_{(n \times (q+1))} \beta_{((q+1) \times p)} + \epsilon_{(n \times p)} \quad (1)$$

dengan  $E(\epsilon_{(i)}) = 0$  dan  $Cov(\epsilon_{(i)}, \epsilon_{(k)}) = \sigma_{ik} I$  dimana  $i, k = 1, 2, \dots, m$  (Johnson dan Wichern, 2002).

Asumsi tambahan yang berkenaan dengan model adalah sebagai berikut :

$$E(Y) = X\beta \text{ atau } E(\epsilon) = 0$$

$cov(y_i) = \Sigma$  untuk semua  $i = 1, 2, \dots, n$  dimana  $y_i^T$  adalah baris ke- $i$  dari matriks  $Y$

$$cov(y_i, y_j) = 0 \text{ untuk semua } i \neq j$$

(Rencher, 2002).

### Koefisien Korelasi Antar Variabel Dependent

Koefisien Korelasi adalah nilai yang menunjukkan kuat atau tidaknya hubungan linier antar dua variabel, cara yang dapat digunakan adalah dengan menghitung matriks korelasi antar variabel (Johnson dan Winchern, 2002).

Dengan persamaan sebagai berikut :

$$r_{ik} = \frac{1}{n-1} \sum_{r=1}^n \left( \frac{x_{ir} - \bar{x}_i}{\sqrt{s_{ii}}} \right) \left( \frac{x_{kr} - \bar{x}_k}{\sqrt{s_{kk}}} \right) \quad (2)$$

dimana  $i = 1, 2, \dots, p$  dan  $k = 1, 2, \dots, p$

nilai  $r_{ik}$  berada antara  $-1 \leq r_{ik} \leq 1$ , ketika  $r_{ik} = 0$  artinya tidak ada hubungan antar variabel, hubungannya sempurna bila  $r_{ik} = \pm 1$ , +1 artinya hubungannya searah dan -1 bila berlawanan arah (Johnson dan Winchern, 2002).

### Pengujian Kebebasan Linier Antar Variabel Dependent

Pengujian kebebasan antar variabel *dependent* digunakan untuk melihat apakah ada hubungan antara tiap variabel *dependent* sebagai syarat pada model regresi multivariat. Jika variabel *dependent* bersifat tidak saling bebas maka analisis regresi multivariat dapat dilanjutkan, jika tidak maka akan dilanjutkan dengan analisis secara univariat.

Untuk menguji kebebasan antar variabel *dependent* dapat dilakukan uji *Barlett Sphericity*.

Urutan pengujiannya sebagai berikut (Nugroho, 2008):

Hipotesis :

$H_0$  : antar variabel *dependent* bersifat saling bebas

$H_1$  : antar variabel *dependent* bersifat tidak saling bebas

Taraf Pengujian :

$\alpha = 0.05$

Kriteria Pengambilan Keputusan :

Uji *Barlett Sphericity* akan menolak

$H_0$  jika nilai  $\lambda_{hitung}^2 > \chi_{\alpha, \frac{p(p-1)}{2}}^2$

Uji *Barlett Sphericity* akan menerima

$H_0$  jika nilai  $\lambda_{hitung}^2 \leq \chi_{\alpha, \frac{p(p-1)}{2}}^2$

Statistik Uji :

$$\lambda_{hitung}^2 = - \left[ (N - 1) - \frac{(2p+5)}{6} \right] \ln |\mathbf{R}| \quad (3)$$

Keterangan :

$N$  : Jumlah observasi

$|\mathbf{R}|$  : Determinan matriks korelasi dari masing-masing variabel *dependent*

$p$  : Jumlah variabel *dependent*

Kesimpulan :

Jika  $H_0$  ditolak berarti antar variabel *dependent* bersifat tidak saling bebas, tetapi jika  $H_0$  diterima berarti antar variabel *dependent* bersifat saling bebas.

### Estimasi Parameter Model Multivariat

Pada regresi multivariat estimator dari kuadrat terkecil  $\hat{\beta}$  dapat dibuat dengan bentuk sebagai berikut:

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

### Akaike Information Criteria Corrected ( $AIC_C$ )

Metode dalam pemilihan model terbaik yang akan digunakan adalah metode  $AIC_C$  yaitu *Akaike Information Criteria Corrected* (Hurvich dan Tsai, 1989).

Berikut perhitungan nilai  $AIC_C$ :

$$AIC_C = n(\ln|\hat{\Sigma}| + q) + 2(p + q + 0.5q(q + 1)) \left( \frac{n}{(n-p-q-1)} \right) \quad (5)$$

dimana  $q$  menyatakan banyaknya variabel *dependent*,  $p$  menyatakan banyaknya variabel *independent* dan  $|\hat{\Sigma}|$  adalah matriks varian kovarian residual berukuran  $N \times N$  (Cavanaugh, 1997).

Himpunan variabel *independent* terbaik  $\mathbf{X}$  adalah himpunan variabel yang memiliki nilai  $AIC_C$  yang terkecil (Cavanaugh, 1997).

### Pengujian Signifikansi Model

Untuk menguji signifikansi model akan dilakukan secara serentak dan juga parsial.

#### Uji Serentak

Pengujian secara serentak dilakukan untuk mengetahui apakah secara keseluruhan parameter signifikan dalam model, dengan urutan pengujiannya sebagai berikut (Johnson dan Wichern, 2002):

Hipotesis :

$H_0$  :  $\beta_{11} = \beta_{12} = \dots = \beta_{q1} = \dots = \beta_{qp} = 0$  (model tidak signifikan)

$H_1$  : Tidak semua  $\beta_{qp}$  sama dengan nol (model signifikan)

Taraf Pengujian :

$\alpha = 0,05$

Kriteria Pengambilan Keputusan :

Jika  $\Lambda_{hitung} \leq \Lambda_{\alpha; p; v_H; v_E}$  maka  $H_0$  ditolak

Jika  $\Lambda_{hitung} > \Lambda_{\alpha; p; v_H; v_E}$  maka  $H_0$  diterima

Statistik Uji :

$$\Lambda = \frac{|E|}{|E+H|} \quad (6)$$

Dengan

$$\mathbf{H} = \hat{\beta}^T \mathbf{X}^T \mathbf{X} \hat{\beta}$$

$$\mathbf{E} = \mathbf{Y}^T \mathbf{Y} - \mathbf{H}$$

$$\mathbf{E} + \mathbf{H} = \mathbf{Y}^T \mathbf{Y}$$

Kesimpulan :

Jika  $H_0$  ditolak maka secara keseluruhan parameter tidak sama dengan nol sehingga model signifikan, tetapi Jika  $H_0$  diterima maka secara keseluruhan parameter sama dengan nol sehingga model tidak signifikan.

#### Uji Parsial

Pengujian ini bertujuan untuk melihat pengaruh signifikan setiap variabel *independent* terhadap variabel-variabel *dependent* secara parsial. Dengan tahapan pengujiannya sebagai berikut (Johnson dan Wichern, 2002):

Hipotesis :

$H_0$  :  $\beta_{q1} = \beta_{q2} = \dots = \beta_{qp} = 0$   
(parameter regresi *independent*  $q$  terhadap variabel *dependent*  $p$ )

tidak berpengaruh secara signifikan)

$H_1$  : Tidak semua  $\beta_{qp}$  sama dengan nol (parameter regresi *independent q* terhadap variabel *dependent p* berpengaruh secara signifikan)

Taraf Pengujian :

$$\alpha = 0,05$$

Kriteria Pengambilan Keputusan :

Jika  $\Lambda_{hitung} \leq \Lambda_{\alpha;p;v_H;v_E}$  maka  $H_0$  ditolak

Jika  $\Lambda_{hitung} > \Lambda_{\alpha;p;v_H;v_E}$  maka  $H_0$  diterima

Statistik Uji :

$$\Lambda = \frac{|E|}{|E+H|} \quad (7)$$

Dengan

$$H = \hat{\beta}^T X^T X \hat{\beta}$$

$$E = Y^T Y - H$$

$$E + H = Y^T Y$$

Kesimpulan :

Jika  $H_0$  ditolak maka parameter regresi *independent q* terhadap variabel *dependent p* berpengaruh secara signifikan, tetapi Jika  $H_0$  diterima maka parameter regresi *independent q* terhadap variabel *dependent p* tidak berpengaruh secara signifikan.

### Keeratan Hubungan Variabel *Dependent* dan *Independent*

Pada regresi multivariat, ukuran yang digunakan untuk mengukur hubungan antar variabel *dependent* dan *independent* adalah Eta Square Lambda yang dinyatakan oleh persamaan

$$\eta_{\Lambda}^2 = 1 - \Lambda \quad (8)$$

dengan  $\Lambda$  adalah nilai wilk's lambda,  $\eta_{\Lambda}^2$  adalah nilai keterkaitan antar variabel *dependent* dan *independent* dengan  $0 \leq \eta_{\Lambda}^2 \leq 1$ . Artinya, semakin mendekati 1 berarti hubungan antara variabel *dependent* dan *independent* semakin erat (Rencher, 2002).

### Uji Asumsi Residual

Ada tiga asumsi residual yang harus dipenuhi dalam model regresi multivariat yaitu residual bersifat identik, residual bersifat saling bebas, dan residual berdistribusi normal.

### Uji Asumsi Residual Bersifat Identik

Untuk menguji syarat ini dapat menggunakan statistik uji *Box's M* (Rencher, 2002)

Hipotesis :

$$H_0 : \Sigma_1 = \Sigma_2 = \dots = \Sigma_k = \dots = \Sigma$$

$$H_1 : \text{Minimal ada satu : } \Sigma_i \neq \Sigma_j \text{ untuk } i \neq j$$

Taraf Pengujian :

$$\alpha = 0,05$$

Kriteria Pengambilan Keputusan :

Jika  $u > \chi_{tabel}^2 = \chi_{\alpha, \frac{1}{2}(k-1)p(p+1)}^2$  maka  $H_0$  ditolak.

Jika  $u \leq \chi_{tabel}^2 = \chi_{\alpha, \frac{1}{2}(k-1)p(p+1)}^2$  maka  $H_0$  diterima.

Statistik Uji :

$$u = -2(1 - c_1) \ln M \quad (9)$$

Dengan

$$\ln M = \frac{1}{2} \sum_{i=1}^k v_i \ln |S_i| -$$

$$\frac{1}{2} (\sum_{i=1}^k v_i) \ln |S_{pool}|$$

$$S_{pool} = \frac{\sum_{i=1}^k v_i S_i}{\sum_{i=1}^k v_i}$$

$$c_1 = \left[ \sum_{i=1}^k \frac{1}{v_i} - \frac{1}{\sum_{i=1}^k v_i} \right] \left[ \frac{2p^2 + 3p - 1}{6(p+1)(k-1)} \right]$$

$$v_i = n_i - 1$$

Keterangan :

$k$  : Banyaknya kelompok

$p$  : Banyaknya variabel residual

$S_i$  : Matriks varian – kovarian residual kelompok ke –  $i$

$n_i$  : Jumlah observasi kelompok ke –  $i$

Kesimpulan :

Jika  $H_0$  ditolak yang berarti matriks-matriks varian kovarian residual adalah tidak homogen, tetapi Jika  $H_0$  diterima yang berarti matriks-matriks varian kovarian residual adalah homogen dan dapat disimpulkan residual bersifat identik.

### Uji Asumsi Residual Bersifat Saling Bebas

Untuk menguji syarat ini dapat dilakukan uji *Barlett Sphericity*. Urutan pengujiannya sebagai berikut (Nugroho, 2008):

Hipotesis :

$H_0$  : Residualnya bersifat saling bebas

$H_1$  : Residualnya bersifat tidak saling bebas

Taraf Pengujian :

$$\alpha = 0,05$$

Kriteria Penolakan :

Uji *Barlett* akan menolak  $H_0$  jika nilai  $\lambda_{hitung}^2 > \chi_{\alpha, \frac{p(p-1)}{2}}^2$

Uji *Barlett* akan menerima  $H_0$  jika nilai  $\lambda_{hitung}^2 \leq \chi_{\alpha, \frac{p(p-1)}{2}}^2$

Statistik Uji :

$$\lambda_{hitung}^2 = - \left[ (N - 1) - \frac{(2p+5)}{6} \right] \ln |R| \quad (10)$$

Keterangan :

$N$  : Jumlah observasi

$|R|$  : Determinan matriks korelasi dari masing-masing variabel residual

$p$  : Jumlah variabel residual

Kesimpulan :

Jika  $H_0$  ditolak berarti residualnya bersifat tidak saling bebas, tetapi Jika  $H_0$  diterima berarti residualnya bersifat saling bebas.

### Uji Asumsi Residual Berdistribusi Normal Multivariat

Urutan pengujiannya sebagai berikut (Johnson dan Wichern, 2002):

Hipotesis :

$H_0$  : Residual berdistribusi normal multivariat

$H_1$  : Residual tidak berdistribusi normal multivariat

Taraf Pengujian :

$$\alpha = 0,05$$

Kriteria Pengambilan Keputusan :

Tolak  $H_0$  jika  $> 50\%$  nilai  $d_i^2 >$

$$\chi_{tabel}^2 = q_{c,3} \left( \frac{i-1}{33} \right)$$

Terima  $H_0$  jika  $\geq 50\%$  nilai  $d_i^2 \leq$

$$\chi_{tabel}^2 = q_{c,3} \left( \frac{i-1}{33} \right)$$

Statistik Uji :

$$d_i^2 = (\hat{\epsilon}_i - \bar{\epsilon})^T S^{-1} (\hat{\epsilon}_i - \bar{\epsilon}) \quad (11)$$

dimana  $i = 1, 2, \dots, n$

Keterangan :

$\hat{\epsilon}_i$  : Vektor residual pengamatan ke  $- i$

$\bar{\epsilon}_i$  : Vektor rata-rata residual pengamatan

$(S_i)^{-1}$  : Invers matriks varian-kovarian residual berukuran  $q \times q$

Kesimpulan :

Jika  $H_0$  ditolak berarti residualnya berdistribusi normal multivariat, tetapi Jika  $H_0$  diterima berarti residualnya berdistribusi normal multivariat.

## HASIL dan PEMBAHASAN

### Statistik Deskriptif Variabel Dependent

Statistik deskriptif variabel *dependent* yang diteliti, bertujuan untuk mengetahui karakteristik dari masing-masing variabel *dependent*.

**Tabel 2.** Statistik Deskriptif Variabel *Dependent*

	N	Minimum	Maximum	Mean
AKB	33	1,9312	3,8198	3,1632
AHH	33	67,7300	78,5900	73,6297
Gizi_Buruk	33	0,0166	1,3250	0,1927
Valid N (listwise)	33			

### Statistik Deskriptif Variabel Independent

Statistik deskriptif variabel *independent* yang diteliti, bertujuan untuk mengetahui karakteristik dari masing-masing variabel *independent*.

**Tabel 3.** Statistik Deskriptif Variabel *Independent*

	N	Min	Max	Mean
Persentase_Tenaga_Medis	33	33,31	99,89	88,82
Persentase_Air_Minum	33	24,00	82,00	61,81
Persentase_Imunisasi	33	66,60	99,70	86,63
Persentase_Jamban	33	50,20	88,40	72,58
Persentase_Diabetes	33	0,80	3,70	2,09
Persentase_ASI	33	25,20	79,70	56,59
Valid N (listwise)	33			

### Pengujian Koefisien Korelasi Antar Variabel Dependent

Koefisien Korelasi adalah nilai yang menunjukkan kuat atau tidaknya hubungan linier antar dua variabel, cara yang dapat digunakan

adalah dengan menghitung matriks korelasi antar variabel (Johnson dan Winchern, 2002).

Dengan menggunakan persamaan (2) diperoleh matriks korelasi sebagai berikut:

$$R = \begin{bmatrix} 1 & -0,51205 & 0,17352 \\ -0,51205 & 1 & -0,43319 \\ 0,17352 & -0,43319 & 1 \end{bmatrix}$$

Berdasarkan matriks **R** di atas dapat di simpulkan karena tidak ada nilai 0 diantara hubungan antar variabel maka hubungan antar variabel *dependent* dikatakan sempurna.

### Pengujian Kebebasan Linier Antar Variabel Dependent

Hipotesis :

$H_0$  : Antar variabel *dependent* bersifat saling bebas

$H_1$  : Antar variabel *dependent* bersifat tidak saling bebas

Taraf Pengujian :

$\alpha = 0,05$

Kriteria Pengambilan Keputusan :

Uji *Barlett Sphericity* akan menolak  $H_0$  jika nilai  $\lambda^2_{hitung} > \chi^2_{\alpha, \frac{p(p-1)}{2}}$

Uji *Barlett Sphericity* akan menerima  $H_0$  jika nilai  $\lambda^2_{hitung} \leq \chi^2_{\alpha, \frac{p(p-1)}{2}}$

Statistik Uji :

Dengan menggunakan persamaan (2.7.1) diperoleh hasil sebagai berikut :

$$\lambda^2_{hitung} = - \left[ (33 - 1) - \frac{(2(3) + 5)}{6} \right] \ln(0,5970) = 15,560$$

Kesimpulan :

Karena nilai  $\lambda^2_{hitung} = 15,560 > \chi^2_{0,05,3} = 7,815$  maka  $H_0$  ditolak yang artinya antar variabel *dependent* bersifat tidak saling bebas.

### Estimasi Parameter

Selanjutnya parameter-parameter pada regresi multivariat perlu diestimasi terlebih dahulu. Estimasi ini bertujuan untuk mendapatkan dugaan besarnya nilai-nilai parameter-parameter pada regresi multivariat. (Rencher, 2002).

Dengan menggunakan persamaan 4 diperoleh matriks  $\hat{\beta}$  berukuran  $(7 \times 3)$  sebagai berikut :

$$\hat{\beta}_{(7 \times 3)} = \begin{bmatrix} -0,000 & 0,000 & -0,000 \\ -0,023 & 0,271 & -0,382 \\ 0,169 & -0,242 & 0,096 \\ -0,491 & 0,015 & -0,259 \\ -0,411 & 0,528 & -0,144 \\ -0,314 & 0,306 & -0,025 \\ -0,022 & -0,027 & 0,090 \end{bmatrix}$$

Matriks ini biasa disebut sebagai parameter regresi.

Dari matriks yang diperoleh di atas dapat dibuat model regresi multivariat sebagai berikut :

$$Y_1 = -0,000 - 0,023X_1 + 0,169X_2 - 0,491X_3 - 0,411X_4 - 0,314X_5 - 0,022X_6$$

$$Y_2 = 0,000 + 0,271X_1 - 0,242X_2 + 0,015X_3 + 0,528X_4 + 0,306X_5 - 0,027X_6$$

$$Y_3 = -0,000 - 0,382X_1 + 0,096X_2 - 0,259X_3 - 0,144X_4 - 0,025X_5 - 0,090X_6$$

### Pemilihan Model Terbaik dengan Metode $AIC_C$

Dengan menggunakan persamaan (5) diperoleh nilai-nilai  $AIC_C$  sebagai berikut :

**Tabel 4.** Nilai  $AIC_C$  dari Kombinasi Variabel **X**

No	Variabel <i>Independent</i>	$AIC_C$
1	$X_1$	73,53876
2	$X_2$	73,53869
3	$X_3$	73,53874
4	$X_4$	73,53873
5	$X_5$	73,53867
6	$X_6$	77,23119
7	$X_1X_2$	70,71974
⋮	⋮	⋮
60	$X_1X_3X_4X_5X_6$	45,13714
61	$X_2X_3X_4X_5X_6$	55,46514
62	$X_1X_2X_3X_4X_5X_6$	33,86527

Berdasarkan nilai dari Tabel 4. di atas didapatkan variabel *independent* terbaik adalah

variabel  $X_1, X_2, X_3, X_4, X_5$  dan  $X_6$  yaitu masing-masing adalah persentase penolong kelahiran dari tenaga medis, persentase rumah tangga menggunakan air minum layak, persentase imunisasi lengkap, persentase rumah tangga menggunakan fasilitas jamban sendiri, persentase penduduk yang menderita penyakit diabetes dan persentase pemberian ASI eksklusif.

### Pengujian Signifikansi Model Regresi

Untuk menguji signifikansi model akan dilakukan secara serentak dan juga parsial.

#### Uji Serentak

Pengujian secara serentak dilakukan untuk mengetahui apakah secara keseluruhan parameter signifikan dalam model, dengan urutan pengujiannya sebagai berikut (Johnson dan Wichern, 2002):

Hipotesis :

$$H_0 : \beta_{11} = \beta_{12} = \dots = \beta_{q1} = \dots = \beta_{qp} = 0 \quad (\text{model tidak signifikan})$$

$$H_1 : \text{Tidak semua } \beta_{qp} \text{ sama dengan nol (model signifikan)}$$

Taraf Pengujian :

$$\alpha = 0,05$$

Kriteria Pengambilan Keputusan :

$$\text{Jika } \Lambda_{hitung} \leq \Lambda_{\alpha;p;v_H;v_E} \text{ maka } H_0 \text{ ditolak}$$

$$\text{Jika } \Lambda_{hitung} > \Lambda_{\alpha;p;v_H;v_E} \text{ maka } H_0 \text{ diterima}$$

Statistik Uji :

Dengan menggunakan persamaan (6) diperoleh hasil sebagai berikut :

$$\Lambda = \frac{|E|}{|E + H|} = \frac{4648,93}{19563,11} = 0,237$$

Kesimpulan :

Karena didapat nilai  $\Lambda_{hitung} = 0,237 \leq \Lambda_{0,05;3;6;30} = 0,392$  maka  $H_0$  ditolak yang artinya secara serentak multivariat, paling tidak ada satu parameter yang signifikan berpengaruh terhadap model.

#### Uji Parsial

Pengujian yang dilakukan setelah menguji model secara serentak yaitu melakukan uji parsial. Pengujian ini bertujuan untuk melihat pengaruh

signifikan setiap variabel *independent* terhadap variabel-variabel *dependent* secara parsial. Tahapan pengujiannya sebagai berikut (Johnson dan Wichern, 2002):

Hipotesis :

a. Variabel  $X_1$  persentase penolong kelahiran oleh tenaga medis  
 $H_0: \beta_{11} = \beta_{12} = \beta_{13} = 0$

$$H_1: \text{Minimal ada satu } \beta_{qp} \neq 0$$

b. Variabel  $X_2$  persentase rumah tangga menggunakan fasilitas air minum sendiri

$$H_0: \beta_{21} = \beta_{22} = \beta_{23} = 0$$

$$H_1: \text{Minimal ada satu } \beta_{qp} \neq 0$$

c. Variabel  $X_3$  persentase imunisasi lengkap

$$H_0: \beta_{31} = \beta_{32} = \beta_{33} = 0$$

$$H_1: \text{Minimal ada satu } \beta_{qp} \neq 0$$

d. Variabel  $X_4$  persentase rumah tangga menggunakan fasilitas jamban sendiri

$$H_0: \beta_{41} = \beta_{42} = \beta_{43} = 0$$

$$H_1: \text{Minimal ada satu } \beta_{qp} \neq 0$$

e. Variabel  $X_5$  persentase penduduk yang menderita penyakit diabetes

$$H_0: \beta_{51} = \beta_{52} = \beta_{53} = 0$$

$$H_1: \text{Minimal ada satu } \beta_{qp} \neq 0$$

f. Persentase  $X_6$  pemberian ASI eksklusif

$$H_0: \beta_{61} = \beta_{62} = \beta_{63} = 0$$

$$H_1: \text{Minimal ada satu } \beta_{qp} \neq 0$$

Taraf Pengujian :

$$\alpha = 0,05$$

Kriteria Pengambilan Keputusan :

Jika  $\Lambda_{hitung} \leq$

$\Lambda_{\alpha;p;v_H;v_E}$  maka  $H_0$  ditolak

Jika  $\Lambda_{hitung} >$

$\Lambda_{\alpha;p;v_H;v_E}$  maka  $H_0$  diterima

Statistik Uji :

Dengan menggunakan persamaan (7) diperoleh hasil yang dibuat dalam bentuk tabel sebagai berikut :

**Tabel 5.** Nilai *Wilk's Lamda* dari setiap variabel *independent*

Variabel	<i>Wilk's Lambda</i>	$\Lambda_{0,05;3;1;31}$
$X_1$	0,680	0,760
$X_2$	0,886	0,760
$X_3$	0,655	0,760
$X_4$	0,601	0,760
$X_5$	0,937	0,760
$X_6$	0,995	0,760

Kesimpulan :

Berdasarkan Tabel 5. di atas didapatkan variabel  $X_1$ ,  $X_3$  dan  $X_4$  yang nilai *Wilk's Lamda* nya  $\leq \Lambda_{0,05;3;1;31} = 0,760$  maka variabel  $X_1$ ,  $X_3$  dan  $X_4$  signifikan terhadap model, yang artinya persentase penolong kelahiran oleh tenaga medis, persentase imunisasi lengkap dan persentase rumah tangga menggunakan jamban sendiri berpengaruh secara signifikan terhadap angka kematian bayi, angka harapan hidup dan juga prevalensi status gizi buruk balita.

### Keeratan Hubungan Variabel *Dependent* dan *Independent*

Pada regresi multivariat, ukuran yang digunakan untuk mengukur hubungan antar variabel *dependent* dan *independent* adalah *Eta Square Lambda*. Dengan menggunakan persamaan (8) diperoleh hasil  $\eta_{\Lambda}^2 = 1 - \Lambda = 1 - 0,2376 = 0,7623$ , karena nilai  $\eta_{\Lambda}^2 = 0,7623$ . Ini dapat dikatakan bahwa model dapat menjelaskan informasi data sebesar 76,23% (Rencher, 2002).

### Pengujian Asumsi Residual

#### Pengujian Asumsi Residual Bersifat Identik

Untuk menguji syarat ini dapat dipergunakan statistik uji *Box's M* (Rencher, 2002)

Hipotesis :

$$H_0 : \Sigma_1 = \Sigma_2 = \dots = \Sigma_k = \dots = \Sigma$$

$$H_1 : \text{Minimal ada satu : } \Sigma_i \neq \Sigma_j \text{ untuk } i \neq j$$

Taraf Pengujian :

$$\alpha = 0,05$$

Kriteria Pengambilan Keputusan :

$$\text{Jika } u > \chi_{tabel}^2 = \chi_{\alpha, \frac{1}{2}(k-1)p(p+1)}^2 \text{ maka } H_0 \text{ ditolak.}$$

$$\text{Jika } u \leq \chi_{tabel}^2 = \chi_{\alpha, \frac{1}{2}(k-1)p(p+1)}^2 \text{ maka } H_0 \text{ diterima.}$$

Statistik Uji :

Dengan menggunakan persamaan (9) diperoleh hasil sebagai berikut :

$$u = -2(1 - 0,0338)(-7,926) = 15,316$$

Kesimpulan :

Karena nilai  $u = 15,316 \leq \chi_{0,05;12}^2 = 21,026$  itu artinya  $H_0$  diterima yang berarti matriks-matriks varian kovarian residual adalah homogen dan dapat disimpulkan residual bersifat identik.

#### Pengujian Asumsi Residual Bersifat Saling Bebas

Untuk menguji syarat ini dapat dilakukan uji *Barlett Sphericity*. Urutan pengujiannya sebagai berikut (Nugroho, 2008):

Hipotesis :

$$H_0 : \text{Antar variabel residualnya bersifat saling bebas}$$

$$H_1 : \text{Antar variabel residualnya bersifat tidak saling bebas}$$

Taraf Pengujian :

$$\alpha = 0,05$$

Kriteria Penolakan :

$$\text{Uji Barlett akan menolak } H_0 \text{ jika nilai } \lambda_{hitung}^2 > \chi_{\alpha, \frac{p(p-1)}{2}}^2$$

$$\text{Uji Barlett akan menerima } H_0 \text{ jika nilai } \lambda_{hitung}^2 \leq \chi_{\alpha, \frac{p(p-1)}{2}}^2$$

Statistik Uji :

Dengan menggunakan persamaan (10) diperoleh hasil sebagai berikut :

$$\lambda_{hitung}^2 = - \left[ (33 - 1) - \frac{(2(3) + 5)}{6} \right] \ln(0,849) = 4,9153$$

Kesimpulan :

Karena nilai  $\lambda_{hitung}^2 = 4,9153 < \chi_{0,05,3}^2 = 7,815$  maka  $H_0$  diterima yang artinya antar residual bersifat *independent* atau saling bebas.

### Pengujian Asumsi Residual Berdistribusi Normal Multivariat

Pengujian asumsi residual berdistribusi normal dapat dilakukan dengan melihat banyaknya nilai  $d_i^2$  yang kurang dari nilai kuantil *Chi-square*. Jika nilai  $d_i^2$  yang kurang dari nilai kuantil *Chi-square* lebih dari 50% maka residual berdistribusi normal multivariat. Urutan pengujiannya sebagai berikut (Johnson dan Wichern, 2002):

Hipotesis :

$H_0$  : Residual berdistribusi normal multivariat

$H_1$  : Residual tidak berdistribusi normal multivariat

Taraf Pengujian :

$$\alpha = 0,05$$

Kriteria Pengambilan Keputusan :

Tolak  $H_0$  jika  $> 50\%$  nilai  $d_i^2 >$

$$\chi_{tabel}^2 = q_{c,3} \left( \frac{i-1}{33} \right)$$

Terima  $H_0$  jika  $\geq 50\%$  nilai  $d_i^2 \leq$

$$\chi_{tabel}^2 = q_{c,3} \left( \frac{i-1}{33} \right)$$

Statistik Uji :

a. Residual  $\hat{\epsilon}_1$

**Tabel 6.** Uji Residual  $\hat{\epsilon}_1$  Berdistribusi Normal

$j$	$d^2_{(j)}$	$q_{c,3} \left( \frac{j-\frac{1}{2}}{33} \right)$
1	0,000	0,045
2	0,000	0,136
⋮	⋮	⋮
33	4,525	2,954

Berdasarkan Tabel 6. sebanyak 88% nilai  $d^2_{(j)} < q_{c,3} \left( \frac{j-\frac{1}{2}}{33} \right)$ .

b. Residual  $\hat{\epsilon}_2$

**Tabel 7.** Uji Residual  $\hat{\epsilon}_2$  Berdistribusi Normal

$j$	$d^2_{(j)}$	$q_{c,3} \left( \frac{j-\frac{1}{2}}{33} \right)$
1	0,002	0,045
2	0,002	0,136
⋮	⋮	⋮
33	6,909	2,954

Berdasarkan Tabel 7. sebanyak 93% nilai  $d^2_{(j)} < q_{c,3} \left( \frac{j-\frac{1}{2}}{33} \right)$ .

c. Residual  $\hat{\epsilon}_3$

**Tabel 8.** Uji Residual  $\hat{\epsilon}_3$  Berdistribusi Normal

$j$	$d^2_{(j)}$	$q_{c,3} \left( \frac{j-\frac{1}{2}}{33} \right)$
1	0,001	0,045
2	0,001	0,136
⋮	⋮	⋮
33	18,541	2,954

Berdasarkan Tabel 8. sebanyak 93% nilai  $d^2_{(j)} < q_{c,3} \left( \frac{j-\frac{1}{2}}{33} \right)$ .

Kesimpulan :

Karena masing-masing variabel residual  $\geq 50\%$  nilai  $d_i^2 \leq \chi_{tabel}^2 = q_{c,3} \left( \frac{i-\frac{1}{2}}{33} \right)$  maka setiap variabel residual berdistribusi normal sehingga secara bersama-sama (multivariat) variabel

tersebut juga dapat dianggap memenuhi asumsi normal multivariat (Nugroho, 2008).

kesehatan di Indonesia dan juga menggunakan metode-metode pemilihan model terbaik yang lain.

## KESIMPULAN dan SARAN

### Kesimpulan

Berdasarkan hasil analisis yang telah dilakukan, diperoleh beberapa kesimpulan sebagai berikut:

1. Faktor-faktor yang mempengaruhi derajat kesehatan yaitu persentase penolong kelahiran yang dilakukan oleh tenaga medis, persentase imunisasi lengkap dan persentase rumah tangga menggunakan fasilitas jamban sendiri.
2. Berdasarkan faktor-faktor yang mempengaruhi derajat kesehatan yang diperoleh, dapat dibuat model regresi multivariat terbaiknya sebagai berikut :

$$Y_1 = -0,00000011 - 0,23233 X_1 - 0,49106 X_3 - 0,4112 X_4$$

$$Y_2 = 0,00000003 + 0,27196 X_1 + 0,015128 X_3 + 0,528617 X_4$$

$$Y_3 = -0,000000078 - 0,38242 X_1 - 0,25989 X_3 - 0,14425 X_4$$

Dari hasil perhitungan nilai *Eta Square Lambda* diperoleh hasil  $\eta_{\Lambda}^2 = 1 - \Lambda = 1 - 0,23 = 0,7623$ , ini dapat dikatakan bahwa model dapat menjelaskan informasi data sebesar 76,23%.

### Saran

Saran yang dapat penulis sampaikan antara lain :

1. Untuk pemerintah dapat lebih memfokuskan pada peningkatan persalinan yang dibantu oleh tenaga medis, meningkatkan pemberian imunisasi secara lengkap, dan juga banyak memberikan penyuluhan kepada masyarakat untuk dapat menggunakan fasilitas jamban sendiri agar dapat menekan angka kematian bayi, menurunkan status gizi buruk balita, serta dapat meningkatkan angka harapan hidup pada masyarakat di Indonesia.
2. Untuk penelitian selanjutnya dapat memasukkan variabel-variabel lain yang diduga mempengaruhi derajat

## DAFTAR PUSTAKA

- Aminuddin., Sudarno., dan Sugito. (2013). *Pemilihan Model Regresi Linier Multivariat Terbaik Dengan Kriteria Mean Square Error*. UNDIP.
- Anonim. (2003). *WHO Definition of Health*. <http://who.int/about/definitor/> diakses pada tanggal 3 Maret 2015.
- Anonim. (2013). *Profil Kesehatan Indonesia Tahun 2012*. Depkes. Jakarta.
- Anonim. (2014a). *Profil Kesehatan Indonesia Tahun 2013*. Depkes. Jakarta.
- Anonim. (2014b). *Badan Pusat Statistik*. <http://Sirusa.bps.go.id/> diakses pada tanggal 12 Desember 2014.
- Anonim. (2015). *Masalah Kesehatan di Indonesia*. <http://www.academia.edu/5273136/> diakses pada tanggal 20 Januari 2015.
- Anton, H. (1995). *Aljabar Linear Elementer*. Erlangga. Jakarta.
- Bilodeau, M., dan Brenner, D. (1999). *Teory of Multivariate Statistics*. Springer-Verlag New York Inc. New York.
- Blum, H. L. (1984). *The Environment of Health*. Human Sciences Press. New York.
- Breiman, L., and Friedman, J. H. (1997). *Predicting Multivariate Response in Multiple Linier Regression*. Journal of the Royal Statistical Society Vol 59, No 1.
- Cavanaugh, J. E. (1997). *Unifying the Derivations for the Akaike and Corrected Akaike Information Criteria*. Departement of Statistics, University of Missouri. Columbia. <http://myweb.uiowa.edu/> diakses pada tanggal 5 Desember 2014.
- Hurvich, C. M., and Tsai, C. L. (1989). *Regression and Time Series Model*

*Selection in Small Samples*. Biometrika  
Vol 76, No 2.

Johnson, R.A., and Wichern, D. (2002). *Applied Multivariate Statistical Analysis*. Prentice Hall. New Jersey.

Lay, D. C. (2006). *Linier Algebra and Its Applications*. Pearson Education Inc. Boston.

Notoatmodjo, S. (2010). *Kesehatan Masyarakat Ilmu dan Seni*. Rieneke Cipta. Jakarta.

Nugroho, S. (2008). *Statistika Multivariat Terapan*. Unib Press. Bengkulu.

Rencher, A.C. (2002). *Methods of Multivariate Analysis*. John Wiley & Sons Inc. New York.

Riskiyanti, R., dan Wulandari. (2010). *Analisis Regresi Multivariat Berdasarkan Faktor-Faktor yang Mempengaruhi Derajat Kesehatan di Provinsi Jawa Timur*. Jurnal Statistika Institut Teknologi Sepuluh November Surabaya.  
<http://digilib.its.ac.id/> diakses pada tanggal 16 November 2014.

Sawer, S. (2010). *Multivariate Linier Models*.  
<http://www.math.wustl.edu/> diakses pada tanggal 25 Desember 2014.

Widarjono, A. (2005). *Ekonometrika Teori dan Aplikasi Untuk Ekonomi dan Bisnis*. Ekonisia Fakultas Ekonomi Universitas Islam Indonesia. Yogyakarta.

# ANALISIS MULTIDIMENSIONAL SCALING PADA POSISI WILAYAH DI PROVINSI BENGKULU BERDASARKAN ATRIBUT PRODUKSI KELAPA SAWIT

Arif Rahman Afrianto<sup>1</sup>, Sigit Nugroho<sup>2</sup>, Fachri Faisal<sup>3</sup>

Mahasiswa Jurusan Matematika FMIPA, Universitas Bengkulu<sup>1</sup>

Dosen Jurusan Matematika FMIPA, Universitas Bengkulu<sup>2,3</sup>

Email : Arifrahman1804@gmail.com<sup>1</sup>

## Abstrak

Penelitian ini mengaplikasikan analisis multidimensional scaling terhadap posisi wilayah di provinsi Bengkulu berdasarkan variabel produksi kelapa sawit yaitu luas areal, jumlah populasi tanaman, tanaman belum menghasilkan, tanaman menghasilkan, tanaman rusak, produktivitas tanaman, produksi, petani, harga tandan buah segar dan curah hujan. Data yang digunakan diperoleh dari badan Pusat Statistik Provinsi Bengkulu dan Dinas Perkebunan Provinsi Bengkulu. Pengolahan data menggunakan MDS menghasilkan nilai Stress 0.00381 dan nilai RSQ 0.99998. Hasil penelitian ini menunjukkan bahwa posisi wilayah yang diperoleh pada setiap objek-objek (Kabupaten/ Kota) yang berada pada perceptual map memiliki tingkat kesesuaian yang besar, setiap objek (kabupaten/ kota) yang berada pada setiap kuadran I, II, III ataupun IV memiliki kemiripan yang dekat berdasarkan produksi kelapa sawit.

**Kata kunci:** analisis multidimensional scaling, multidimensional scaling metrik, Perceptual Map, kelapa sawit.

## 1. PENDAHULUAN

Kelapa Sawit adalah salah satu dari tanaman palma penghasil minyak, tanaman kelapa sawit sangat cocok dibudidayakan sebagai salah satu cabang usaha baik perorangan, kelompok maupun perusahaan [1]. Tanaman Kelapa Sawit atau dalam bahasa latinnya dikenal dengan *Elaeis Guineensis* Jacq ini berasal dari Nigeria, Afrika Barat. Tanaman ini merupakan tanaman perkebunan yang dominan di Indonesia [3].

Menurut UU No 18 Tahun 2004 Pasal 1 disebutkan bahwa perkebunan adalah segala kegiatan yang mengusahakan pada tanah dan/atau media tumbuh lainnya dalam ekosistem yang sesuai, mengolah dan memasarkan barang dan jasa hasil tanaman tersebut, dengan bantuan ilmu pengetahuan dan teknologi, pemodal dan manajemen untuk mewujudkan kesejahteraan bagi pelaku usaha perkebunan dan masyarakat. Tanaman Kelapa Sawit merupakan tanaman industri andalan bagi Indonesia karena masih bertahan pada saat terjadinya krisis ekonomi berkepanjangan dan merupakan

salah satu komoditas perkebunan yang menyumbang devisa besar bagi negara [2]. Berdasarkan data dari Dinas Perkebunan Provinsi Bengkulu dapat disimpulkan bahwa ada dua faktor yang mempengaruhi produksi kelapa sawit yaitu faktor internal dan faktor eksternal, faktor internal itu sendiri terdiri dari beberapa atribut yaitu luas area, jumlah populasi tanaman, tanaman belum menghasilkan, tanaman menghasilkan, tanaman rusak, produktivitas tanaman, dan produksi sedangkan untuk faktor eksternal terdiri dari curah hujan, harga Tandan Buah Segar dan jumlah petani.

Sejalan dengan uraian diatas, faktor internal dan eksternal merupakan atribut yang dapat digunakan untuk mengetahui posisi Kabupaten/ Kota di Wilayah Provinsi Bengkulu, salah satu metode yang dapat digunakan adalah Multidimensional Scaling. Dalam menyampaikan suatu data atau informasi, seringkali akan lebih mudah dan menarik menyampaikannya dalam bentuk gambar, termasuk dalam menampilkan data-data (atribut) suatu objek [4].

Multidimensional Scaling adalah salah satu teknik peubah ganda yang dapat digunakan untuk menentukan posisi suatu objek dengan objek lainnya berdasarkan penilaian kemiripannya [5]. Metode ini dianggap dapat mengetahui posisi Kabupaten/ Kota di Wilayah Provinsi Bengkulu berdasarkan atribut produksi kelapa sawit dan faktor-faktor yang mempengaruhi produksi kelapa sawit, sehingga dari atribut tersebut dapat diketahui Kabupaten/ Kota mana yang memiliki kemiripan atribut pada produksi kelapa sawit. Dari uraian di atas dapat dirumuskan permasalahan yaitu : Bagaimana Posisi Kabupaten/Kota di Wilayah Provinsi Bengkulu berdasarkan atribut produksi kelapa sawit dengan menggunakan Analisis Multidimensional Scaling ? Sehingga terpenuhi tujuan untuk mengetahui Posisi Kabupaten/ Kota di Wilayah Provinsi Bengkulu berdasarkan atribut produksi Kelapa Sawit dengan menggunakan Analisis Multidimensional Scaling.

## **2. METODE PENELITIAN**

### **2.1 Jenis Penelitian**

Penelitian ini merupakan penelitian statistik terapan, yaitu suatu penelitian yang dilakukan dengan mengaplikasikan metode statistika ke dalam bidang-bidang tertentu. Pada penelitian ini, analisis Multidimensional Scaling digunakan untuk menganalisis posisi wilayah di Provinsi Bengkulu berdasarkan atribut produksi kelapa sawit serta program yang digunakan pada penelitian ini yaitu program SPSS 16.

#### **a. Pengambilan Data**

Data yang digunakan dalam penelitian ini merupakan data skunder, yang di ambil dari BPS Provinsi Bengkulu dan Dinas Perkebunan Provinsi Bengkulu.

## 2.2 Identifikasi Variabel

Identifikasi variabel dilakukan untuk menentukan variabel-variabel yang akan diukur dalam penelitian.

Variabel yang digunakan dalam penelitian ini	Objek yang diamati pada penelitian ini
X1 = Luas Area (Ha)	BU = Bengkulu Utara
X2 = Jumlah Populasi Tanaman (Ha)	MK = Mukomuko
X3 = TBM (Tanaman Belum Menghasilkan) (Ha)	RL = Rejang lebong
X4 = TM (Tanaman Menghasilkan)	KP = Kepahiang
X5 = TR (Tanaman Rusak) (Ha)	LB = Lebong
X6 = Produktivitas Tanaman (Kg/Ha)	BS = Bengkulu Selatan
X7 = Produksi (Ton)	SL = Seluma
X8 = Petani (KK)	KR = Kaur
X9 = Harga Tandan Buah Segar (Rp)	BT = Bengkulu Tengah
X10 = Curah Hujan (Mm)	KB = Kota Bengkulu

## 2.3 Teknik Analisis Data

Tahapan-tahapan yang akan dilakukan dalam penelitian ini adalah :

### 1. Mendeskripsikan Data

Mencari data skunder yang akan digunakan, Data yang digunakan dalam penelitian ini yaitu data dari Badan Pusat Statistik (BPS) dan Dinas Perkebunan tahun 2014

### 2. Melakukan analisis menggunakan metode MDS

Mencari jarak Euclid masing-masing wilayah

Melakukan analisis Multidimensional Scaling dari jarak Euclid yang telah diperoleh

### 3. Menginterpretasikan peta persepsi yang telah diperoleh dari analisis Multidimensional Scaling.

## 3. HASIL DAN PEMBAHASAN

### 3.1 Deskripsi Data

Data yang digunakan dalam penelitian ini adalah data sekunder, yang diperoleh dari Badan Pusat Statistik Provinsi Bengkulu dan Dinas Perkebunan Provinsi Bengkulu yaitu data produksi tanaman kelapa sawit di Provinsi Bengkulu. Variabel data yang digunakan dalam penelitian ini terdiri dari 10 variabel produksi tanaman kelapa sawit yaitu Luas Area, Jumlah Populasi Tanaman, TBM (Tanaman Belum Menghasilkan), TM (Tanaman Menghasilkan), TR (Tanaman Rusak), Produktivitas Tanaman, Produksi,

Petani, Harga Tandan Buah Segar, dan Curah Hujan. Objek data yang digunakan dalam penelitian ini yaitu seluruh Kabupaten/ Kota di Provinsi Bengkulu yang terdiri dari 10 Kabupaten/ Kota.

### 3.2 Analisis Multidimensional Scaling

Setelah melakukan analisis deskriptif data, langkah selanjutnya yaitu melakukan analisis Multidimensional Scaling dimana akan digunakan program SPSS 16.0 untuk pengolahan dan analisis posisi wilayah Kabupaten/ Kota di Provinsi Bengkulu. Analisis posisi wilayah ini adalah studi kasus yang dilakukan untuk mengetahui posisi wilayah Kabupaten/ Kota melalui tingkat kemiripan (Similarity) dalam bentuk koordinat titik-titik objeknya yang akan digambarkan dalam dimensi ganda, posisi tersebut memperlihatkan kedekatan (Proximity) antar objeknya.

Data persepsi kemiripan adalah berupa data produksi tanaman kelapa sawit yaitu atribut-atribut yang dianggap mempengaruhi produksi kelapa sawit, yang dapat dilihat pada Tabel 4.1. Kemudian dari data tersebut diolah menggunakan program SPSS 16.0 sehingga diperoleh jarak Euclid pada masing-masing Kabupaten/ Kota yang dapat dilihat pada Tabel 4.2. Berdasarkan jarak antar tiap pasang Kabupaten/ Kota, hal ini menunjukkan kemiripan masing-masing wilayah Kabupaten/ Kota terhadap atribut produksi kelapa sawit. Apabila semakin dekat jarak antara dua objek, maka semakin mirip pula objek tersebut terhadap atribut produksi kelapa sawit.

**Tabel 4.1** Data Atribut Produksi Kelapa Sawit dan Objek Produksi Kelapa Sawit

Atribut	OBJEK									
	BU	MK	RL	KP	LB	BS	SL	KR	BT	KB
X1	4324,60	4036,70	1639,98	665,00	1921,82	1186,10	2400,44	2369,05	1223,94	151,70
X2	27848	98714	486	131	493	14093	31381	8114	7282	1282
X3	18089	71474	200	91	75	11688	20910	4000	3680	1568
X4	8978	26032	283	32	416	2392	10288	4096	3602	241
X5	781	1208	3	8	2	13	183	18	0	12
X6	63748	267872	466	222	236	34318	66702	12360	13780	6352
X7	3524	3748	2330	2440	3147	2936	3190	3090	3745	4051
X8	17162	43873	237	88	327	12627	13645	6709	6159	1081
X9	1150	1150	1150	1150	1150	1150	1150	1150	1150	1150
X10	276,91	276,91	276,92	276,91	276,91	276,91	276,91	276,91	276,91	276,91

**Tabel 4.2** Jarak Euclid Pasangan Objek (Kabupaten/ Kota)

	BU	MK	RL	KP	LB	BS	SL	KR	BT	KB
BU	0.000	224,818.121	73787.571	74259.181	73958.250	34220.198	6886.405	58020.246	57337.798	68016.273
MK	224,818.121	0.000	298011.764	298426.379	298216.252	258510.971	220738.451	282691.460	281874.803	292036.813
RL	73787.571	298011.764	0.000	1115.099	917.266	40269.281	77793.563	16481.204	16846.951	6560.503
KP	74259.181	298426.379	1115.099	0.000	1553.542	40685.116	78244.264	17029.961	17304.724	6706.153
LB	73958.250	298216.252	917.266	1553.542	0.000	40462.130	77980.667	16581.712	16956.831	6693.757
BS	34220.198	258510.971	40269.281	40685.116	40462.130	0.000	38698.760	24826.541	24005.882	34480.207
SL	6886.405	220738.451	77793.563	78244.264	77980.667	38698.760	0.000	62183.933	61488.555	2258.368
KR	58020.246	282691.460	16481.204	17029.961	16581.712	24826.541	62183.933	0.000	72019.512	11877.096
BT	57337.798	281874.803	16846.951	17304.724	16956.831	24005.882	61488.555	2258.368	0.000	11574.156
KB	68016.273	292036.813	6560.503	6706.153	6693.757	34480.207	72019.512	11877.096	11574.156	0.000

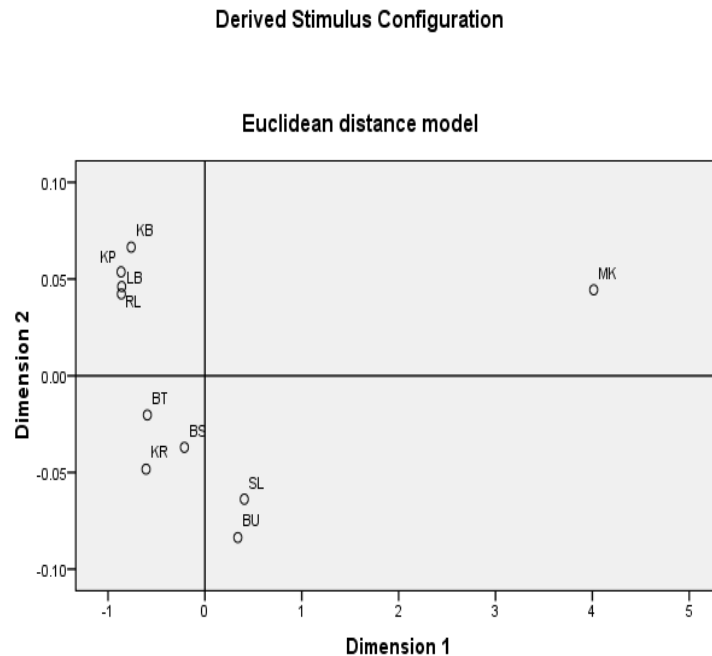
Setelah mengetahui jarak Euclid selanjutnya diperoleh titik-titik stimulus atau koordinat maka akan dilakukan pemetaan yang bertujuan untuk mengetahui posisi wilayah Kabupaten/ Kota di Provinsi Bengkulu berdasarkan atribut produksi kelapa sawit. Koordinat stimulusnya dapat dilihat pada Tabel 4.3

**Tabel 4.3 Koordinat Stimulus**

Stimulus	Dimensi	
	1	2
BU	0.3403	-0.0837
MK	4.0141	0.0444
RL	-0.8589	0.0461
KP	-0.8657	0.0537
LB	-0.8622	0.0423
BS	-0.2116	-0.0370
SL	0.4070	-0.0638
KR	-0.6074	-0.0483
BT	-0.5948	-0.0203
KB	-0.7611	0.0665

Setelah diperoleh titik-titik stimulusnya, selanjutnya juga diperoleh peta posisi yaitu posisi wilayah Kabupaten/ Kota di Provinsi Bengkulu terhadap atribut produksi kelapa sawit. Analisis Multidimensional Scaling Menghasilkan konfigurasi peta yang dapat dilihat dari nilai Stress dan nilai RSQ yang dicapai. Apabila nilai Stress yang diperoleh sebesar 0 mengindikasikan bahwa konfigurasi yang dihasilkan sudah sangat sempurna. Artinya semakin kecil nilai Stress maka akan semakin baik jarak antar objek dalam peta yang

dihasilkan. Seiring dengan hal itu, jika nilai RSQ lebih besar dari 0.5 atau mendekati 1 maka model tersebut juga sudah sangat sempurna. Peta Posisi Wilayah di Provinsi Bengkulu berdasarkan atribut produksi kelapa sawit tersebut dapat dilihat pada Gambar 4.1.



Gambar 4.1 Peta Posisi untuk Objek (Kabupaten/ Kota).

Pada Gambar 4.1 terlihat bahwa KB (Kota Bengkulu), LB (Kabupaten Lebong), RL (Kabupaten Rejang Lebong) dan KP (Kepahiang) memiliki kemiripan dalam posisi wilayah Kabupaten/ Kota terhadap atribut produksi kelapa sawit karena terletak pada kuadran yang sama yaitu kuadran II. Kemudian untuk BT (Kabupaten Bengkulu Tengah), KR (Kabupaten Kaur) dan BS (Kabupaten Bengkulu Selatan) memiliki kemiripan dalam posisi wilayah Kabupaten/ Kota terhadap atribut produksi kelapa sawit yaitu terletak pada kuadran III, untuk SL (Kabupaten Seluma) dan BU (Kabupaten Bengkulu Utara) memiliki kemiripan dalam posisi wilayah Kabupaten/ Kota terhadap atribut produksi karena terletak pada kuadran yang sama yaitu kuadran IV. MK (Kabupaten Mukomuko) memiliki perbedaan yang cukup besar dibandingkan dengan Objek (Kabupaten) lainnya karena berada paling jauh dari Kabupaten lainnya yang ada di Provinsi Bengkulu yaitu terletak pada kuadran I. Nilai Stress yang diperoleh untuk analisis Multidimensional Scaling pada Posisi Wilayah Kabupaten/ Kota di Provinsi Bengkulu terhadap atribut produksi adalah sebesar 0.00381 sedangkan untuk nilai RSQ yang diperoleh adalah sebesar 0.99998 yang berarti bahwa konfigurasi dan model Multidimensional Scaling yang diperoleh sangat baik sesuai dengan Tabel 2.1 dan model Multidimensional Scaling yang diperoleh berupa peta posisi seperti Gambar 4.1.

Hal ini menunjukkan bahwa penggambaran koordinat pada tiap objek Kabupaten/ Kota dalam Perceptual Map telah memiliki tingkat kesesuaian yang tinggi dengan data jarak antar tiap objek Kabupaten/ Kota. Dalam kasus ini ditetapkan dimensi yang akan digunakan yaitu 2 dimensi, karena yang dibutuhkan dalam dimensi ini hanya dua dimensi saja. Untuk dimensi yang lebih tinggi akan mengalami kesulitan dalam menggambarkan peta dan menginterpretasikan peta. Peta posisi diatas menunjukkan posisi Wilayah Kabupaten/ Kota di Provinsi Bengkulu berdasarkan atribut produksi kelapa sawit. Dari peta posisi tersebut dapat diketahui jarak yang menginterpretasikan tingkat kemiripan antar Kabupaten/ Kota di Provinsi Bengkulu. Untuk lebih lengkapnya dapat dilihat pada lampiran 2.

#### **4. KESIMPULAN DAN SARAN**

Setelah dilakukan analisis dan pengolahan data, maka pada bab ini di ambil kesimpulan dan diberikan saran untuk penelitian selanjutnya.

##### **4.1 Kesimpulan**

Berdasarkan hasil dan pembahasan yang telah di uraikan pada bab sebelumnya, maka diperoleh informasi sebagai berikut:

Berdasarkan data dari variabel-variabel produksi kelapa sawit yaitu diantaranya Luas Areal, Jumlah Populasi Tanaman, TBM, TM, TR, Produktivitas Tanaman, Produksi, Curahan Tenaga Kerja, Harga Tandan Buah Segar dan Curah Hujan maka posisi wilayah yang diperoleh menginterpretasikan bahwa Objek-objek (Kabupaten/ Kota) yang berada pada kuadran yang sama memiliki kemiripan produksi tanaman kelapa sawit.

1. KB (Kota Bengkulu), LB (Kabupaten Lebong), RL (Kabupaten Rejang Lebong) dan KP (Kepahiang) memiliki kemiripan karena terletak pada satu kuadran yaitu kuadran II maka dipersepsikan memiliki kemiripan dalam posisi wilayah Kabupaten/ Kota terhadap atribut produksi kelapa sawit.
2. BT (Kabupaten Bengkulu Tengah), KR (Kabupaten Kaur) dan BS (Kabupaten Bengkulu Selatan) berada pada kuadran III maka dipersepsikan memiliki kemiripan dalam posisi wilayah Kabupaten/ Kota terhadap atribut produksi kelapa sawit.
3. SL (Kabupaten Seluma) dan BU (Kabupaten Bengkulu Utara) karena terletak pada kuadran yang sama yaitu kuadran IV sehingga dapat dipersepsikan bahwa SL dan BU memiliki kemiripan dalam posisi wilayah Kabupaten/ Kota terhadap atribut produksi kelapa sawit.

4. Sedangkan MK (Kabupaten Mukomuko) memiliki perbedaan yang cukup besar dibandingkan dengan Objek (Kabupaten) lainnya karena berada paling jauh dari Kabupaten lainnya yang ada di Provinsi Bengkulu, MK (Kabupaten Mukomuko) terletak pada kuadran I.

#### 4.2 Saran

Berdasarkan hasil penelitian dan kesimpulan yang diperoleh, penulis dapat memberikan saran sebagai berikut :

1. Dalam penggunaan MDS yang melibatkan atribut perlu diperhatikan apakah atribut-atribut yang terpilih telah secara lengkap menjelaskan seluruh variasi atau perbedaan antar objek sehingga diperoleh dimensi yang relevan dengan posisi objek yang dihasilkan pada peta.
2. Perlu dilakukan kajian lebih lanjut untuk mengetahui faktor-faktor apa saja yang sangat mempengaruhi produksi kelapa sawit.

### 5. DAFTAR PUSTAKA

- [1]. Anonim, 2009, Peluang Investasi Sub Sektor Perkebunan di Provinsi Bengkulu, Dinas Perkebunan Provinsi Bengkulu.
- [2]. Hannum, J., Chairani, H. dan J. Ginting, 2014, Kadar N, P Daun dan Produksi Kelapa Sawit Melalui Penempatan TKKS pada Rorak, Jurnal Online Agroekoteknologi, No. 4, Vol. 2, 1279-1286.
- [3]. Hartono, B. Adiwirman dan G. ME. Manurung, 2014, Teknik Budidaya Tanaman Kelapa Sawit (*Elaeis guineensis* Jacq) Belum Menghasilkan di Lahan Pasang Surut yang Dilakukan Petani Di Kecamatan Bangko Pusako Kabupaten Rokan Hilir, Jom Faperta, No. 2, Vol. 1.  
<http://jom.unri.ac.id/index.php/JOMFAPERTA/article/view/3634/3527>
- [4]. Mattjik, A. A. dan I. M. Sumertajaya, 2011, Sidik Peubah Ganda Dengan Menggunakan SAS, Pertama, IPB PRESS.
- [5]. Walundongo, A. G., Paendong, M. dan T. Manurung, 2014, Penggunaan Analisis Multidimensional Scaling Untuk Mengetahui Kemiripan Rumah Makan di Manado Town Square Berdasarkan Karakteristik Pelanggan, JDC, No. 1, Vol. 3, 30-35.

# PEMODELAN JUMLAH KASUS ANGKA KEMATIAN BAYI (AKB) MENURUT KECAMATAN DI KABUPATEN BENGKULU SELATAN DENGAN *GEOGRAPHICALLY WEIGHTED POISSON REGRESSION* (GWPR)

Intan Purnama Sari<sup>1</sup>, Sigit Nugroho<sup>2</sup>, dan Fachri Faisal<sup>3</sup>  
Program Studi Matematika  
Fakultas Matematika dan Ilmu Pengetahuan Alam  
Universitas Bengkulu

## ABSTRAK

Pada penelitian ini akan digunakan metode *Geographically Weighted Poisson Regression* (GWPR) sebagai statistik untuk menganalisis data spasial dari proses non stasioner. GWPR adalah bentuk lokal dari regresi *Poisson* dimana lokasi pengambilan data diperhatikan yang berasumsi bahwa data berdistribusi *Poisson*. Analisis regresi *Poisson* juga merupakan metode regresi yang digunakan untuk menganalisis data yang variabel dependennya berupa data diskrit. Pada regresi *Poisson* terdapat asumsi yang harus dipenuhi, yaitu nilai varians dan rata-rata dari variabel dependen tersebut sama. Data yang digunakan pada penelitian ini kasus angka kematian bayi dan faktor-faktor yang mempengaruhi di Kabupaten Bengkulu Selatan. Untuk mendapatkan model terbaik maka sejumlah model harus dievaluasi dan nilai AIC untuk setiap model harus dibandingkan. Model yang terbaik adalah model dengan nilai AIC terendah. Berdasarkan nilai AIC antara model regresi *Poisson* dan model GWPR, diketahui model GWPR dengan pembobot fungsi kernel gaussian merupakan model yang lebih baik digunakan untuk menganalisis jumlah kasus angka kematian bayi di Kabupaten Bengkulu Selatan.

**Kata Kunci:** *Geographically Weighted Poisson Regression*, Regresi *Poisson*, Angka Kematian Bayi.

## A. PENDAHULUAN

Angka kematian bayi (AKB) merupakan indikator yang sangat berguna untuk mengetahui status kesehatan anak dan kondisi ekonomi penduduk suatu wilayah dan untuk melihat target penurunan angka kematian bayi. Salah satu cara yang dapat dilakukan untuk menurunkan angka kematian bayi adalah dengan mengetahui faktor-faktor penyebabnya (Elyna, 2012).

Negara Indonesia masih harus berjuang keras untuk memperbaiki indikator pembangunan kesehatan, khususnya angka kematian bayi, karena tren angka kematian bayi selama beberapa tahun terakhir belum menurun. Untuk itu pemerintah harus berupaya keras melalui berbagai program untuk menekan angka kematian bayi (Aulele, 2012).

Metode analisis regresi *Poisson*, dimana analisis regresi merupakan analisis yang

memanfaatkan dua atau lebih variabel sehingga salah satu variabel dapat diramalkan dari variabel lainnya. Pada analisis regresi terdiri dari dua jenis variabel yaitu variabel bebas (variabel independen) dan variabel tak bebas (variabel dependen). Analisis regresi *Poisson* adalah suatu model yang digunakan untuk menganalisis hubungan antara variabel dependen yang berdistribusi *Poisson* dengan variabel-variabel bebasnya (Amalia, 2012).

Analisis regresi *Poisson* juga merupakan metode regresi yang digunakan untuk menganalisis data yang variabel dependennya berupa data diskrit. Pada regresi *Poisson* terdapat asumsi yang harus dipenuhi, yaitu nilai varians dan rata-rata dari variabel dependen tersebut sama.

Beberapa variabel independen berpengaruh secara global, sedangkan yang lainnya dapat mempertahankan pengaruh spasialnya. Hubungan jumlah kematian bayi dengan faktor-faktor penyebabnya dapat diketahui dengan

menggunakan analisis regresi *Poisson*, karena jumlah kematian bayi merupakan data cacah (count data) dengan peluang kejadian kecil yang tergantung pada interval waktu tertentu atau suatu wilayah tertentu (Afri, 2013).

Pada penelitian ini akan di bahas mengenai deskripsi kasus AKB dan faktor-faktor yang mempengaruhi melalui peta tematik, menyusun model regresi *Poisson* dan model GWPR untuk mendapatkan faktor-faktor yang mempengaruhi jumlah kasus AKB, serta membandingkan model regresi *Poisson* dan model GWPR untuk mendapatkan model yang terbaik pada pemodelan kasus AKB menurut Kecamatan di Kabupaten Bengkulu Selatan, dengan menggunakan pembobot fungsi kernel gaussian dan kriteria pemilihan model terbaik AIC (*Akaike's Information Criterion*).

## B. KAJIAN PUSTAKA

### 1. Analisis Regresi

Regresi merupakan suatu teknik dalam statistika untuk menentukan suatu persamaan garis atau kurva dengan cara meminimumkan penyimpangan atau deviasi antara data pengamatan dan nilai-nilai dugaannya. Regresi digunakan untuk menduga nilai-nilai suatu variabel dependen dari nilai variabel (peubah) yang sudah diketahui atau diasumsikan ada hubungan dengannya (Nugroho, 2008).

Analisis regresi adalah metode untuk menemukan hubungan antara variabel dependen yang dinotasikan dengan  $y$  dan variabel independen yang biasanya dinotasikan dengan  $x_1, x_2, \dots, x_n$ . Apabila variabel independen lebih dari satu maka digunakan analisis linier berganda.

### 2. Regresi *Poisson*

Regresi *Poisson* merupakan suatu bentuk analisis regresi yang digunakan untuk memodelkan data yang berbentuk diskrit (count data), misalnya data tersebut dilambangkan dengan  $Y$  yaitu banyaknya kejadian yang terjadi dalam suatu periode waktu dan/atau wilayah tertentu dimana  $Y$  merupakan variabel dependen dan  $X$  merupakan variabel independen (Notje, 2010).

Model regresi *Poisson* merupakan *Generalized Linier Model* (GLM) yang data dependen diasumsikan berdistribusi *Poisson*. Nugroho (2008), suatu peubah acak diskrit  $Y$  dikatakan memiliki sebaran *Poisson* dengan parameter  $\mu > 0$  jika memiliki fungsi kepekatan peluang seperti berikut:

$$f(y, \mu) = \frac{e^{-\mu} \mu^y}{y!}; y = 0, 1, 2, \dots \quad (1)$$

Selanjutnya untuk membangun model regresi *Poisson*, dimisalkan sampel acak  $Y_i \sim \text{Poisson}(\mu_i)$ ,  $i = 1, 2, \dots, n$  dan rata-rata  $\mu_i$  bergantung pada vektor peubah bebas (peubah penjelas)  $x_i$  dan vektor koefisien regresi  $\beta$ , yaitu:

$$\mu_i = (x_i, \beta) = x_i^T \beta \quad (2)$$

Di mana:

$$x_i = [1 \ x_{1i} \ x_{2i} \ \dots \ x_{ki}]^T;$$

$$\beta = [\beta_0 \ \beta_1 \ \beta_2 \ \dots \ \beta_k]^T.$$

Estimasi parameter-parameter ( $\mu$ ) pada model regresi *Poisson* menggunakan metode MLE (*Maximum Likelihood Estimation*). Metode estimasi maksimum likelihood dapat dilakukan jika distribusi data diketahui. Langkah pertama yang dilakukan adalah menentukan fungsi *likelihood* dari model regresi *Poisson*. Dengan mengamsusikan  $X_1, X_2, \dots, X_n$  adalah variabel random yang *mutually independent*.

Pengujian parameter model regresi *Poisson* dapat menggunakan ukuran *Goodness of Fit* yang disebut devians (*deviance*). Devians merupakan perbandingan antara nilai maksimum *likelihood* untuk model yang lebih lengkap  $L(\beta)$  dengan nilai maksimum *likelihood* untuk model yang kurang lengkap  $L(\lambda)$ , bentuk persamaan sebagai berikut:

$$D(\hat{\beta}) = -2 \ln \left( \frac{L(\hat{\beta})}{L(\hat{\lambda})} \right) \quad (3)$$

### 3. Model *Geographically Weighted Poisson Regression*.

GWPR merupakan suatu metode statistika yang sebenarnya yang pengembangan dari regresi *Poisson* namun yang membedakan dalam metode ini yaitu memperhatikan pembobot berupa letak lintang dan letak bujur dari titik-titik pengamatan yang diamati dan disimbolkan  $(u_i, v_i)$ . Model GWPR menghasilkan penaksir parameter model yang bersifat lokal untuk setiap titik pengamatan.

Berikut ini adalah model *Geographically Weighted Poisson Regression* (GWPR) dimana  $u_i$  sebagai koordinat lintang dan  $v_i$  sebagai koordinat bujur yang digunakan sebagai pembobot penaksiran parameter nantinya (Amalia, 2012)

$$y_i \sim \text{poisson} \left[ \exp \left( \sum_{k=0}^p \beta_k(u_i, v_i) x_{ik} \right) \right] \quad (4)$$

Penaksiran parameter dilakukan dengan menggunakan metode kemungkinan maksimum. Langkah awal dari metode GWPR adalah dengan membentuk fungsi *likelihood*.

Berdasarkan bentuk *likelihood* diatas kemudian dilakukan operasi logaritma natural sebagai berikut:

$$\ln L(\boldsymbol{\beta}) = \sum_{i=1}^n (-\mu(\mathbf{x}_i, \boldsymbol{\beta}) + y_i \ln \mu(\mathbf{x}_i, \boldsymbol{\beta}) - \ln y_i!) \quad (5)$$

Nilai estimasi diperoleh dengan memaksimumkan bentuk diferensial untuk fungsi *log likelihood* tersebut persamaan di atas dibuat sama dengan nol, sehingga diperoleh:

$$\frac{\partial \ln L^*(\boldsymbol{\beta}(u_i, v_i))}{\partial \boldsymbol{\beta}^T(u_j, v_j)} = \sum_{i=1}^n [y_i \mathbf{x}_j - \mathbf{x}_j \exp(\mathbf{x}_j^T \boldsymbol{\beta}(u_j, v_j))] \quad (6)$$

$$W_{ij}(u_i, v_i) = 0$$

Karena fungsi pada persamaan (6) berbentuk implisit, maka digunakan suatu prosedur iterasi numerik yaitu metode *Newton-Raphson*. Secara umum persamaan untuk iterasi *Newton-Raphson* adalah:

$$\boldsymbol{\beta}_{m+1}(u_i, v_i) = \boldsymbol{\beta}_m(u_i, v_i) - \mathbf{H}_{(m)}^{-1}(\boldsymbol{\beta}_m(u_i, v_i)) \mathbf{g}_{(m)}(\boldsymbol{\beta}_m(u_i, v_i)) \quad (7)$$

Pengujian persamaan model GWPR dengan model *Poisson* digunakan untuk mengetahui apakah ada perbedaan pengujian regresi *Poisson* dengan GWPR dengan perhitungan nilai F. Menguji persamaan antara model GWPR dengan model regresi *Poisson* dengan hipotesis berikut:

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

$$H_1: \beta_j(u_i, v_i) \neq \beta_j$$

Statistik uji:

$$F_{hit} = \frac{\text{Devians Model A} / df_A}{\text{Devians Model B} / df_B} \quad (8)$$

Tolak  $H_0$  jika nilai  $F_{hit} > F_{(\alpha, df_A, df_B)}$ , berarti terdapat perbedaan yang signifikan antara model regresi *Poisson* dan model GWPR.

#### 4. Penentuan Bandwidth

*Bandwidth* adalah ukuran jarak fungsi pembobot dan sejauh mana pengaruh lokasi terhadap lokasi lain. *Bandwidth* dapat dianalogikan sebagai radius dari suatu lingkaran, sehingga sebuah titik yang berada di dalam radius lingkaran masih dianggap memiliki pengaruh. Proses

penaksiran parameter model GWPR di suatu titik  $(u_i, v_i)$  membutuhkan pembobot spasial dimana pembobot yang digunakan adalah fungsi kernel gauss:

$$w_{ij}(u_i, v_i) = \exp\left(-\left(\frac{d_{ij}}{b}\right)^2\right) \quad (9)$$

Di mana :

$w_{ij}$  : nilai bobot dari observasi pada lokasi k-j untuk penaksiran koefisien pada lokasi ke-i.

$d_{ij}$  : jarak *Euclidean* antara i dan j dengan

$$d_{ij} = \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2}$$

antara lokasi  $(u_i, v_i)$  ke lokasi  $(u_j, v_j)$ .

$b$  : ukuran *bandwidth* yang ditetapkan didefinisikan dengan ukuran matriks jarak.

$b_{i(p)}$ : ukuran *bandwidth* yang diadaptasi sebagai jarak tetangga terdekat ke-i dimana  $i=1, 2, \dots, n$ .

Salah satu metode yang digunakan untuk memilih *bandwidth* optimum adalah metode *Cross Validation* (CV) yang didefinisikan sebagai berikut:

$$CV(b) = \sum_{i=1}^n (y_i - \hat{y}_{\neq i}(b))^2 \quad (10)$$

Metode yang digunakan untuk memilih model terbaik untuk GWPR yaitu *Akaike Information Criterion* (AIC) yang didefinisikan sebagai berikut:

$$AIC = D(b) + 2K(b) \quad (11)$$

Model terbaik adalah dengan nilai AIC terkecil.

#### 5. Angka Kematian Bayi

Angka kematian bayi adalah salah satu indikator penting dalam menentukan tingkat kesehatan masyarakat. Keberhasilan pembangunan di suatu wilayah juga dapat dilihat dari angka kematian bayi dan angka harapan hidup (Aulele, 2012).

Kegunaan Angka Kematian Bayi untuk pengembangan perencanaan berbeda antara kematian neo-natal dan kematian bayi yang lain. Kematian neo-natal disebabkan oleh faktor endogen yang berhubungan dengan kehamilan maka program-program untuk mengurangi angka kematian neo-natal adalah yang bersangkutan dengan program pelayanan

kesehatan Ibu hamil, misalnya program pemberian pil besi dan suntikan anti tetanus. Sedangkan Angka Kematian Post-NeoNatal dan Angka Kematian Anak serta Kematian Balita dapat berguna untuk mengembangkan program imunisasi, serta program-program pencegahan penyakit menular terutama pada anak-anak, program penerangan tentang gizi dan pemberian makanan sehat untuk anak dibawah usia 5 tahun (Hajarisman,2013).

## 6. Faktor-faktor yang mempengaruhi Angka Kematian Bayi

Kondisi geografis, sosial budaya dan ekonomi yang berbeda antara wilayah yang satu dengan wilayah yang lain menyebabkan terjadinya keragaman spasial. Sehingga faktor-faktor yang berpengaruh terhadap jumlah kematian bayi berbeda antara satu wilayah dengan wilayah lain. Beberapa penyebab kematian bayi dikarenakan berat badan lahir rendah, asfiksia, tetanus, infeksi dan masalah pemberian minuman.

## C. METODE PENELITIAN

Data yang digunakan pada penelitian ini merupakan data sekunder yang diperoleh dari Buku Profil Dinas Kesehatan Bengkulu Selatan dan Badan Pusat Statistik (BPS). Pada penelitian ini mencakup data AKB dan faktor-faktor yang mempengaruhi AKB pada tahun 2013 di Kabupaten Bengkulu Selatan.

Tahap analisis yang digunakan untuk mencapai tujuan penelitian adalah sebagai berikut :

1. Untuk mengetahui deskripsi dari Kecamatan di Kabupaten Bengkulu Selatan berdasarkan variabel penelitian, maka pendeskripsian menggunakan peta tematik Kabupaten Bengkulu Selatan yaitu mendeskripsikan variabel dependen ( $Y$ ) dan variabel independen ( $X$ ) dari sudut kewilayahannya yang dibuat dari program ArcMap 10.2.
2. Melakukan analisis statistik deskriptif jumlah kasus AKB di Kabupaten Bengkulu Selatan.
3. Menganalisis model Regresi *Poisson*. Model regresi *Poisson* merupakan *Generalized Linear Model* (GLM) dengan data responnya (komponen random) diasumsikan berdistribusi *Poisson*. Pada model regresi *Poisson*, biasanya *link function* yang digunakan adalah log sehingga  $\ln(\mu_i) = \eta_i$  dan fungsi hubungan untuk model regresi *Poisson* mempunyai logaritma.

- a. Penaksiran parameter model regresi *Poisson* dengan metode MLE.
  - b. Menguji signifikansi parameter model regresi *Poisson* secara serentak dan parsial. Uji signifikansi parameter secara serentak menggunakan nilai devians sebagai statistik uji. Uji signifikansi parsial menggunakan nilai  $T_{hitung}$  yang dibandingkan dengan  $T_{tabel}$ .
  - c. Menghitung nilai AIC model regresi *Poisson*.
4. Menganalisis model *Geographically Weighted Poisson Regression*. Model GWPR menghasilkan penaksiran parameter model yang bersifat lokal untuk setiap titik lokasi pengamatan. Langkah berikutnya adalah menghitung jarak *Euclidean* antara titik pengamatan dan menghitung matriks pembobot yang dilanjutkan dengan penaksiran parameter dan uji kesesuaian model.
    - a. Menghitung jarak *euclidean* antar lokasi pengamatan berdasarkan posisi geografis.
    - b. Menentukan *bandwidth* optimum berdasarkan kriteria CV minimum.
    - c. Menghitung matriks pembobot dengan menggunakan fungsi kernel gaussian.
    - d. Menaksir parameter model GWPR dengan metode MLE.
    - e. Melakukan pengujian model GWPR untuk menguji signifikansi dari faktor geografis. Model regresi GWPR secara parsial pada wilayah (Kecamatan) menggunakan nilai devians sehingga didapatkan variabel apa saja yang berpengaruh terhadap wilayah yang telah diestimasi modelnya. Kemudian hal tersebut dilakukan untuk tiap wilayah Kecamatan di Kabupaten Bengkulu Selatan.
    - f. Menghitung nilai AIC model GWPR.
    - g. Pengujian kesesuaian model antara model regresi *Poisson* dengan model GWPR menggunakan nilai devians dari masing-masing model dan dicari nilai  $F_{hitung}$ .
  5. Mendapatkan model terbaik pada pemodelan jumlah kasus AKB di Kabupaten Bengkulu Selatan dengan Regresi *Poisson* dan GWPR menggunakan pembobot fungsi kernel gaussian. Perbandingan model Regresi *Poisson* dengan model GWPR menggunakan nilai AIC, model yang baik adalah model yang memiliki nilai AIC paling kecil.

## D. HASIL DAN PEMBAHASAN

Untuk melihat karakteristik dari masing-masing variabel maka disajikan statistika deskriptif yang dapat dilihat pada tabel sebagai berikut:

Tabel 1. Statistika Deskriptif

Variabel	Rata-rata	Standar Deviansi	Min	Mak
(Y)	1,79	1,89	0,00	6,00
(X <sub>1</sub> )	99,27	6,75	86,52	112,31
(X <sub>2</sub> )	1,03	8,35	87,26	117,77
(X <sub>3</sub> )	88,06	16,49	43,00	113,63
(X <sub>4</sub> )	12,15	9,61	0,00	27,90
(X <sub>5</sub> )	54,58	14,01	23,30	87,60
(X <sub>6</sub> )	0,68	0,23	0,32	1,12

Pada pendeskripsian setiap variabel penelitian, variabel penelitian dikategorikan menjadi 3 kelompok tingkatan, yaitu tingkat tinggi, sedang dan rendah.

### Pemodelan Angka Kematian Bayi dengan menggunakan Regresi *Poisson*

Data yang digunakan dalam penerapan model regresi *Poisson* yaitu data Angka Kematian Bayi yang di ambil dari 11 kecamatan dan 14 puskesmas di Kabupaten Bengkulu Selatan pada tahun 2013. Peluang *Poisson* adalah sangat kecil dan ukuran contoh belum tentu diketahui. Uji kesesuaian distribusi *Poisson* dapat dilakukan dengan Uji *Kolmogorov-Smirnov Z* dimana hipotesis pengujianya sebagai berikut:

$H_0$ : Data berdistribusi *Poisson*

$H_1$ : Data tidak berdistribusi *Poisson*

$H_0$  diterima karena nilai *Asymp. Sig. (2-tailed)* = 0,696 >  $\alpha = 0,05$  artinya bahwa data angka kematian bayi berdistribusi *Poisson*.

Langkah awal untuk analisis model GWPR dengan membentuk model regresi *Poisson* terlebih dahulu. Sebelum membentuk regresi *Poisson* maka perlu dilakukan uji *kolinieritas* untuk mengetahui apakah variabel independen telah memenuhi kondisi tidak saling berkorelasi. uji *kolinieritas* dengan melihat nilai *VIF* diperoleh variabel independen menunjukkan nilai kurang dari 10 menunjukkan bahwa variabel independen di Kabupaten Bengkulu Selatan dapat dikatakan tidak saling berkorelasi. Nilai *Pearson's Chi-Square* model dibagi derajat bebasnya adalah 1.593. Nilai tersebut lebih besar dari satu yang berarti dapat

disimpulkan bahwa data mengalami *overdispersi*. Kasus *overdispersi* akan menyebabkan model yang terbentuk menjadi estimasi parameter yang bias.

Adanya salah satu aspek spasial yaitu sifat heterogenitas spasial adalah sebagai syarat bisa dilakukan pemodelan data dengan menggunakan pendekatan titik dengan regresi *Poisson*. Untuk mendeteksi ada atau tidaknya heterogenitas spasial dalam model dilakukan uji *Breusch-pagan* hasil dari  $BP = 0 < \chi_p^2 = 3,84$  kesimpulan  $H_0$  di terima, Data tidak normal dan independen homogenitas atau tidak terjadi heterogenitas.

Pengujian parameter model regresi *Poisson* menggunakan ukuran *Goodness of Fit* yang melihat nilai devians (*deviance*), yang bertujuan untuk mengetahui apakah ada tidaknya pengaruh variabel independen secara bersama-sama terhadap variabel dependen. Karena nilai  $D(\hat{\beta}) > \chi_{\alpha,k}^2$  maka  $H_0$  di tolak, artinya paling sedikit ada satu  $\beta_k \neq 0$  yang berpengaruh terhadap model regresi *Poisson* dengan menggunakan taraf 5%.

Pengujian Parameter secara parsial digunakan untuk mengetahui parameter mana saja yang memberikan pengaruh signifikasi terhadap model. Persentase bayi BBLR ( $X_4$ ) merupakan variabel independen yang signifikan mempengaruhi variabel dependen karena nilai  $w = 8.110 > \chi_{(\alpha,1)}^2 = 3,84$ .

### Model *Geographically Weighted Poisson Regression*

Langkah-langkah untuk membangun model ini adalah dengan memilih *bandwidth* (G) optimum pemodelan GWPR, selain metode *Cross Validation* kriteria lain yang dapat digunakan yaitu dengan *AIC* minimum (Nakaya, 2004).

Pembobot fungsi kernel gaussian dan nilai *bandwidth* sebesar 103,870 dengan nilai *AIC* minimum 23,777. Nilai *bandwidth* digunakan untuk menentukan matriks pembobot untuk setiap wilayah ke-*i*. Analisis GWPR adalah dengan menentukan letak geografis dari masing-masing wilayah atau titik pengamatan. Sebelum menghitung matriks pembobot yang harus dilakukan yaitu mencari jarak *euclidean* ( $d_{ij}$ ) antar kecamatan.

Estimasi parameter model GWPR dengan memasukkan pembobot spasial ke dalam perhitungan dengan menggunakan metode *Maximum Likelihood Estimation*.

Pengujian parameter secara parsial dilakukan untuk mengetahui parameter mana saja yang berpengaruh signifikan terhadap variabel dependen pada tiap-tiap lokasi dengan hipotesis sebagai berikut:

$$H_0: \beta_j(u_i, v_i) = 0$$

$$H_1: \beta_j(u_i, v_i) \neq 0 ; j=1,2, \dots, n.$$

Hasil pengujian model GWPR sama dengan hasil pengujian model regresi Poisson secara parsial untuk setiap kecamatan di Kabupaten Bengkulu Selatan, dapat disimpulkan variabel yang signifikan pada taraf  $\alpha = 0,05$  untuk setiap kecamatan di Bengkulu Selatan hanya variabel persentase bayi BBLR ( $X_4$ ) = 2,85.

Pengujian persamaan model GWPR dengan model *Poisson* digunakan untuk mengetahui apakah ada perbedaan pengujian regresi *Poisson* dengan GWPR dengan perhitungan nilai  $F_{hitung}$ .  $F_{hitung}$  yaitu  $1,000 < F_{(0,05;7;7)} = 3,79$  apabila digunakan tingkat signifikansi  $\alpha = 0,05$ , maka diperoleh keputusan  $H_0$  diterima berarti tidak terdapat perbedaan yang signifikan antara model GWPR dan model regresi *Poisson* dengan fungsi *kernel gaussian*.

Model GWPR tidak lebih baik dari model regresi *Poisson* karena pada kasus angka kematian bayi menurut kecamatan di Kabupaten Bengkulu Selatan tidak mampu menjelaskan variansi dari kejadian angka kematian bayi di Kabupaten Bengkulu Selatan. Hal ini dikarenakan asumsi homoskedastisitas tidak dilanggar dengan kata lain tidak terdapat pengaruh spasial atau tidak terdapat faktor geografis yang berpengaruh terhadap kasus angka kematian bayi menurut kecamatan di Kabupaten Bengkulu Selatan.

### **Perbandingan Model Regresi *Poisson* dan Model *Geographically Weighted Poisson Regression*.**

Perbandingan antara model regresi *Poisson* dengan model GWPR baik dengan menggunakan pembobot fungsi kernel gaussian dilakukan untuk mengetahui model mana yang lebih baik diterapkan untuk jumlah kematian bayi menurut kecamatan di Kabupaten Bengkulu Selatan. Kriteria kebaikan model yang digunakan adalah dengan membandingkan nilai AIC dari kedua model tersebut.

Nilai AIC dari model GWPR sebesar 29,975 sedangkan nilai AIC dari model regresi *Poisson* sebesar 48,667. Hasil perbandingan nilai AIC menunjukkan bahwa nilai AIC model GWPR lebih

kecil dibandingkan nilai AIC model regresi *Poisson*. Namun hal ini tetap disimpulkan bahwa Model Regresi *Poisson* lebih baik dari pada Model GWPR.

## **E. KESIMPULAN DAN SARAN**

### **Kesimpulan**

Dari hasil penelitian data dan pembahasan pada BAB 4 dapat diperoleh kesimpulan sebagai berikut :

1. Berdasarkan jumlah angka kematian bayi beserta faktor-faktor yang mempengaruhinya, rata-rata jumlah kasus angka kematian bayi menurut kecamatan di Kabupaten Bengkulu Selatan pada tahun 2013 adalah 1,79%. Kejadian angka kematian bayi di kabupaten Bengkulu Selatan menunjukkan pola menyebar, dapat dilihat dari pemetaan setiap variabel dan model angka kematian bayi untuk tiap Kecamatan di Kabupaten Bengkulu Selatan yang juga bersifat mengelompok, sehingga memiliki varians secara spasial.
2. Berdasarkan model regresi *Poisson* diperoleh faktor-faktor yang signifikan mempengaruhi angka kematian bayi yaitu persentase bayi BBLR ( $X_4$ ). Sehingga model regresi *Poisson* adalah sebagai berikut:  

$$\mu_i = \exp(-0,472 + 0,069 X_4).$$
3. Model regresi *Poisson* lebih baik dari pada model GWPR dengan melihat nilai uji asumsi. Model regresi *Poisson* lebih baik dikarenakan asumsi homoskedastisitas tidak dilanggar. Perbandingan model regresi *Poisson* dan model GWPR dengan melihat nilai AIC, model GWPR nilai AIC lebih kecil dari model regresi *Poisson*.

### **Saran**

Saran yang diberikan penulis kepada Dinas Kesehatan Kabupaten Bengkulu Selatan adalah setelah dilakukan penelitian Dinas Kesehatan di Kabupaten Bengkulu Selatan yaitu lebih memperhatikan faktor-faktor penyebab angka kematian bayi per kecamatan di Kabupaten Bengkulu Selatan dari hasil yang telah dicapai pada penelitian ini. Penelitian selanjutnya mengenai model regresi *Poisson*, hendaknya digunakan regresi *Poisson* yang asumsi homogenitas errornya terpenuhi sehingga model GWPR dapat digunakan.

## F. DAFTAR PUSTAKA

- Afri, L.E. 2013. Model Regresi Binomial Negatif Terboboti Geografis untuk Data Kematian Bayi. Program Studi Pendidikan Matematika. Fakultas Keguruan dan Ilmu Pendidikan. Universitas Pasir Pengaraian.
- Amalia, R dan Puhadi. 2012. *Pemodelan Jumlah Balita Gizi Buruk di Jawa Timur dengan Geographically weighted Poisson Regression*. Jurnal sains dan seni. Jurusan statistik. ITS. Surabaya.
- Anonim, 2014. Badan Pusat Statistik (2014), Kabupaten Bengkulu Selatan dalam Angka Tahun 2014. Bengkulu Selatan.
- Anonim. 2014. *Profil Kesehatan Provinsi Bengkulu Tahun 2013*. Bengkulu.
- Aulele, S.N. 2012. *Pemodelan Kematian Bayi di Provinsi Maluku Tahun 2010 dengan menggunakan Regresi Poisson*. Staf Jurusan Matematika, FMIPA, Unpati. Yogyakarta.
- Fotheringham, A. S., C Brunsdon and M Charlton. 2002. *Geographically Weighted Regression the analysis of spatially varying relationship*. University of Newcastle, UK.
- Hanjarisma, N. 2013. *Pendugaan Angka Kematian Bayi melalui Model Regresi Poisson Bayes Berhierarchy Dua-Level*. Program Studi Statistika, UIB. Bandung.
- Nakaya, T., A.S Fotheringham., C Brunsdon and M Charlton. 2005. *Geographically Weighted Poisson Regression for Disease Association Mapping, Statistics in Medicine*. Ritsumeikan University. Japan.
- Notje, S.A. 2010. *Model Geographically Weighted Poisson Regression. Studi Kasus : Jumlah Kematian Bayi di Jawa Timur & Jawa Tengah tahun 2007*. Surabaya: Program Pascasarjana. Institut Teknologi Sepuluh Nopember.
- Nugroho, S. 2008. *Dasar-Dasar Metode Statistika*. Bengkulu : PT. Gramedia Widiasarana Indonesia.
- Nugroho, S. 2008. *Pengantar Statistika Matematika Edisi Pertama*. Universitas Bengkulu. Bengkulu.

# PEMODELAN KEJADIAN POISSON PADA *SMALL AREA ESTIMATION* MENGGUNAKAN METODE *HIERARCHICAL BAYES* DENGAN FUNGSI PRIOR LOG-NORMAL

Anita Nurul Fitriani\*, Sigit Nugroho\*\*, Etis Sunandi\*\*\*  
\*Mahasiswa Matematika, \*\*Dosen Pembimbing Utama,  
\*\*\*Dosen Pembimbing Pendamping

Tujuan penelitian ini adalah menjelaskan dan mempelajari langkah-langkah pendugaan parameter pada *Small Area Estimation* menggunakan metode *Hierarchical Bayes* dengan fungsi prior Log-Normal. Di dalam penelitian ini terdapat salah satu metode untuk menduga parameter pada *Small Area Estimation* yaitu *Hierarchical Bayes*. *Small Area* didefinisikan sebagai bagian dari populasi yang memiliki ukuran contoh kecil. Metode *Hierarchical Bayes* ini digunakan pada data cacahan dengan fungsi prior Log-Normal. Metode pada penelitian ini adalah studi literatur. Hasil penelitian menunjukkan bahwa parameter untuk data cacahan pada *Small Area Estimation* yaitu pendugaan rata-rata dari fungsi prior Log-Normal memberikan hasil pendugaan dengan ketelitian yang lebih baik dibandingkan pendugaan langsung.

**Kata Kunci:** *Hierarchical Bayes*, *Small Area Estimation*, fungsi prior Log-Normal

## I. PENDAHULUAN

### 1.1 Latar Belakang

Suatu survei dilakukan dengan tujuan untuk menduga parameter populasi yang cakupannya meliputi suatu area yang luas. Tidak hanya dimanfaatkan untuk menduga parameter keseluruhan, hasil survei juga digunakan untuk menduga parameter pada bagian dari populasi (*sub population*) atau area yang lebih kecil (*small area*), misalnya informasi pada level provinsi, kabupaten, bahkan mungkin level kecamatan. Jika ukuran sampel yang tersedia dari survei tersebut sudah cukup memenuhi atau sesuai dengan keakuratan yang dispesifikasikan untuk menduga parameter *sub population*, maka pendugaan parameter dapat dilakukan secara langsung (*direct estimation*). Pendugaan langsung umumnya didasarkan pada penarikan contohnya (*sampling technique*). Misalnya, *simple random sampling*, *systematic random sampling*, *stratified random sampling*, *cluster random sampling*, dan lain sebagainya.

Menurut Kurnia (2009) permasalahan penting yang ditimbulkan oleh metode pendugaan langsung terhadap area kecil tersebut ada dua. Pertama, penduga yang dihasilkan merupakan penduga tak bias tetapi memiliki ragam yang besar karena diperoleh dari ukuran contoh yang kecil. Kedua, apabila pada suatu area kecil ke-*i* tidak terwakili di dalam survei, maka tidak memungkinkan dilakukan pendugaan secara langsung. Oleh karena itu, perlu dikembangkan suatu metode pendugaan dengan cara tidak langsung (*indirect estimation*), dengan tujuan

untuk meningkatkan keefektifan ukuran contoh dan menurunkan keragaman sehingga lebih akurat. Pendugaan tersebut dikenal sebagai *Small Area Estimation* (SAE).

*Small Area Estimation* (SAE) merupakan suatu teknik statistika untuk menduga parameter-parameter subpopulasi yang ukuran contohnya kecil. Pengertian *small area* tidak hanya mengacu pada suatu area/wilayah (mengacu pada pengertian secara geografis), namun *small area* dapat juga mengacu pada suatu bagian dari populasi yang didefinisikan menurut kriteria tertentu, misalnya berdasarkan jenis kelamin, ras, usia dan lain-lain (Rao, 2003).

Berdasarkan sumber data yang digunakan, SAE diklasifikasikan menjadi, (1) data *cross section* yaitu data yang terdiri dari satu objek namun memerlukan sub objek-sub objek lainnya yang berkaitan atau yang berada dalam objek induk tersebut pada suatu waktu, (2) data *time series* yaitu data yang terdiri dari satu objek namun terdiri dari beberapa periode waktu, seperti harian, mingguan, bulanan, dan tahunan, (3) gabungan antara data *cross section* dan data *time series*. Berdasarkan tipe penarikan kesimpulan yang digunakan, SAE dapat digolongkan sebagai, (1) metode berdasarkan rancangan (*design-based*), (2) metode berdasarkan model (*model-based*) yang terdiri atas pendekatan *frequentist* dan Bayesian, (3) gabungan antara *design-based* dan *model-based*. Berdasarkan tingkat (*level*) ketersediaan data penyerta, model SAE terdiri atas, (1) area level dan (2) unit level (Handayani, 2015).

Metode SAE telah banyak digunakan berbasis model (*model-based area estimation*) yang diperkenalkan dalam Ghosh dan Rao (1994), diantaranya adalah *Best Linear Unbiased Prediction (BLUP)*, *Empirical Best Linear Unbiased Prediction (EBLUP)*, *Empirical Bayes (EB)*, dan *Hierarchical Bayes (HB)*. Metode *BLUP* dan *EBLUP* merupakan metode untuk data kontinu sedangkan *EB* dan *HB* adalah metode untuk data biner atau cacahan. Pemodelan SAE yang sedang banyak dikaji adalah pemodelan dengan menerapkan kaidah *Bayes (Bayesian Estimation)*, karena lebih menguntungkan dengan mempunyai nilai *Mean Square Error (MSE)* yang lebih kecil dibandingkan *BLUP* atau *EBLUP*.

Konsep *Bayes* digunakan karena pendugaan pada domain dengan sampel yang kecil sangat membutuhkan informasi pendukung, baik yang berasal dari penelitian sebelumnya, bahkan dari sebuah penilaian yang subjektif atau spesifik tiap domain/area. Pada pemodelan SAE dimasukkan pula pengaruh acak area yang sangat memungkinkan untuk menganggap bahwa parameter yang tidak diketahui dalam model adalah bersifat acak/random mengikuti distribusi tertentu serta dipengaruhi oleh serangkaian variabel prediktor yang dianggap sebagai informasi prior.

Metode estimasi Bayes yang akhir-akhir ini banyak dikembangkan adalah *EB* dan *HB*. Salah satu kelebihan kedua metode ini adalah masalah inferensi parameternya yang dapat dijelaskan dengan menggunakan fungsi kepekatan posterior dugaan. Tetapi, pada metode *EB* hanya memasukkan informasi prior empiris (berdasarkan data), sedangkan metode *HB* memasukkan informasi prior empiris dan prior subjektif. Selain itu Kelebihan metode *HB* adalah masalah inferensinya relatif lebih jelas dan tepat serta komputasinya juga relatif lebih mudah dengan menggunakan teknik *Marcov Chain Monte Carlo (MCMC)* dan juga mampu mengatasi model dengan efek acak mengikuti distribusi selain distribusi Normal (Hajarisman, 2013).

Dari beberapa metode yang telah dikemukakan, pemilihan model yang tepat memungkinkan akan menghasilkan estimasi yang cukup baik dari sampel yang kecil, asalkan tersedia data pendukung yang lengkap dan berkaitan dengan data yang hendak diteliti. Penerapan pendugaan parameter pada data cacahan adalah menggunakan distribusi Poisson,

karena menyatakan banyaknya “kejadian langka” yang terjadi dalam suatu selang waktu tertentu.

Terdapat banyak penelitian mengenai *Small Area Estimation* dengan metode bayes di indonesia, diantaranya Kismiantini (2007) meneliti “Pendekatan Bayes Empirik pada Pendugaan Statistik Area Kecil Berbasis Model Poisson-Gamma dengan Peubah Penyerta”. Nadhiroh (2008) meneliti “*Zero Inflated Negative Binomial Models In Small Area Estimation*”. Hadi dan Nusyirwan (2008) meneliti “Penduga Maksimum Likelihood untuk Parameter Dispersi Model Poisson-Gamma dalam Konteks Penduga Area Kecil”. Kismiantini (2008) meneliti “Pendugaan Berbasis Model untuk Kasus Biner pada *Small Area Estimation*”. Kismiantini (2010) meneliti “Penerapan Metode Bayes Empirik pada Pendugaan Area Kecil untuk Kasus Biner”. Sunandi (2013) meneliti “Model Spasial Bayes dalam Pendugaan Area Kecil dengan Peubah Respon Biner”. Rumiati (2012) meneliti “*Empirical Bayesian Method For The Estimation Of Literacy Rate At Sub-District Level Case Study: Sumenep District Of East Java Province*”. Noviyanti (2014) meneliti “Pendekatan *Small Area Estimation* pada *Scan Statistic* untuk Pendeteksian Kantong Kemiskinan”.

Berdasarkan beberapa penelitian diatas, model yang digunakan adalah model Logit-Normal dan Poisson-Gamma, oleh karena itu penulis tertarik untuk meneliti **Pemodelan Kejadian Poisson Pada *Small Area Estimation* Menggunakan Metode *Hierarchical Bayes* Dengan Fungsi Prior Log-Normal.**

## 1.2 Rumusan Masalah

Adapun rumusan masalah yang akan dibahas dalam penelitian ini yaitu :

1. Bagaimana langkah-langkah pemodelan kejadian Poisson pada *small area estimation* menggunakan metode *hierarchical bayes* dengan fungsi prior Log-Normal ?
2. Bagaimana penerapan pemodelan kejadian Poisson pada *small area estimation* menggunakan metode *hierarchical bayes* dengan fungsi prior Log-Normal ?

## 1.3 Batasan Masalah

Penulisan proposal skripsi ini terbatas pada pembahasan mengenai :

1. Data yang digunakan berdistribusi Poisson.
2. Fungsi prior yang digunakan berdistribusi Log-Normal.

## 1.4 Tujuan Penelitian

Tujuan yang ingin dicapai dalam penelitian ini yaitu :

1. Untuk memberikan langkah-langkah pemodelan kejadian Poisson pada *small area estimation* menggunakan metode *hierarchical bayes* dengan fungsi prior Log-Normal.
2. Untuk mengaplikasikan teladan penerapan mengenai pemodelan kejadian Poisson pada *small area estimation* menggunakan metode *hierarchical bayes* dengan fungsi prior Log-Normal.

## 1.5 Manfaat Penelitian

Manfaat yang ingin dicapai dari hasil penelitian ini yaitu dapat memberikan tambahan wawasan mengenai *Small Area Estimation* untuk data cacahan dengan menggunakan metode *Hierarchical Bayes (HB)* dan pengaplikasiannya pada fungsi prior Log-Normal ke dalam teladan penerapan.

# II. METODE PENELITIAN

## 2.1 Jenis Penelitian

Penelitian ini merupakan studi literatur yang berupa membandingkan pendugaan parameter secara langsung dengan pendugaan parameter kejadian Poisson pada *small area estimation* menggunakan metode *hierarchical bayes* dengan fungsi prior Log-Normal.

## 2.2 Waktu

Waktu yang direncanakan dalam penelitian dan penyusunan skripsi ini selama kurang lebih empat bulan.

## 2.3 Tempat

Penyusunan skripsi ini dikerjakan dan diselesaikan di FMIPA Universitas Bengkulu, sedangkan data yang digunakan dalam penelitian ini diperoleh dengan cara pembangkitan data simulasi pada bahasa pemrograman R-3.2.2.

## 2.4 Data Penelitian

Adapun data-data yang diperlukan dalam penelitian ini adalah sebagai berikut :

- a. Data banyaknya sampel pada area kecil ke- $i$  yang dibangkitkan pada program R-3.2.2.
- b. Data kejadian Poisson pada area kecil ke- $i$  yang dibangkitkan pada program R-3.2.2.

## 2.5 Metode Analisis Data

Penelitian ini dianalisis menggunakan metode *Hierarchical Bayes (HB)* dengan langkah sebagai berikut :

1. Membangkitkan data banyaknya sampel dan kejadian Poisson untuk setiap area kecil ke- $i$  pada program R-3.2.2.

2. Menentukan parameter rata-rata ( $\theta_i$ ) dengan

langkah sebagai berikut :

- a. Menghitung  $\theta_i$  dengan rumus (2.6)

sehingga varian dapat ditentukan.

- b. Menghitung MSE dari varian dan  $\theta_i$  yang

telah didapatkan.

- c. Menghitung RMSE

3. Pendugaan tidak langsung rata-rata ( $\theta_i$ )

dengan metode *Hierarchical Bayes (HB)* yaitu prosedur MCMC. Prosedur MCMC tersebut adalah :

- a. Menghitung nilai  $\hat{\mu} \sim N\left(\frac{1}{m} \sum_i \xi_i, \frac{\sigma^2}{m}\right)$

- b. Menghitung nilai

$$\hat{\sigma}^2 \sim G\left(\frac{m}{2} + a, \frac{1}{2} \sum_i (\xi_i - \mu)^2 + b\right)$$

- c. Menghitung nilai  $\hat{\theta}_i^{HB}$

- d. Menghitung  $var(\hat{\theta}_i^{HB})$  dan bias

- e. Menghitung penduga  $MSE(\hat{\theta}_i^{HB})$  dan

$$RMSE(\hat{\theta}_i^{HB})$$

4. Menganalisis dan membandingkan nilai  $RMSE$  hasil perhitungan parameter rata-rata

( $\theta_i$ ) yang diperoleh pada metode

*Hierarchical Bayes (HB)* dan pendugaan langsung.

# III. PEMODELAN KEJADIAN POISSON PADA SMALL AREA ESTIMATION MENGGUNAKAN METODE HIERARCHICAL BAYES DENGAN FUNGSI PRIOR LOG-NORMAL DAN LANGKAH-LANGKAH PENDUGAAN PARAMETER

### 3.1 Metode Hierarchical Bayes dengan Fungsi Prior Log-Normal

Menurut Rao (2003) pendugaan parameter menggunakan metode *Hierarchical Bayes (HB)* dengan prior Log-Normal didefinisikan dengan model sebagai berikut :

- i.  $y_i | \theta_i \sim iid \text{Poisson}(e_i \theta_i)$
  - ii.  $\xi_i = \log(\theta_i) | \mu, \sigma^2 \sim iid N(\mu, \sigma^2)$
  - iii.  $f(\mu, \sigma^2) \propto f(\mu) f(\sigma^2)$  ( $\mu$  dan  $\sigma^2$  saling bebas) dengan  $f(\mu) \propto 1$ ;  $\frac{1}{\sigma^2} \sim G(a, b)$ ;  $a \geq 0, b > 0$ .
- $$(4.1)$$

Model (i) merupakan distribusi sampel variabel  $y$  pada data. Parameter  $\theta_i$  adalah rata-rata yang akan ditaksir pada penelitian ini dengan menggunakan fungsi log pada model (ii). Untuk melakukan penaksiran rata-rata ( $\theta_i$ ) perlu ditaksir terlebih dahulu parameter  $\hat{\mu}$  dan  $\hat{\sigma}^2$ . Distribusi prior yang tepat untuk  $\hat{\sigma}^2$  adalah *Invers Gamma* dengan parameter  $a$  dan  $b$  yang ditentukan (Maiti, 1998).

Rata-rata pada model *Hierarchical Bayes* akan ditaksir sebagai mean  $E(\theta_i | y)$  dengan varian  $V(\theta_i | y)$  dari distribusi posterior bersama :

$$f(\theta_i | y) = \int_{\hat{\mu}} \int_{\hat{\sigma}^2} f(\theta_i, \hat{\mu}, \hat{\sigma}^2 | y) d\hat{\sigma}^2 d\hat{\mu} \quad (4.2)$$

Sebaran posterior dari bentuk (4.2) memiliki bentuk *opened form*, maka solusi alternatif yang dapat digunakan adalah dengan menghitung besaran posterior melalui integrasi numerik. Menurut Hajarisman (2013), salah satu metode yang dapat digunakan adalah *Markov Chain Monte Carlo (MCMC)*. Tujuan dari MCMC adalah membangun suatu peluang rantai markov hingga pada akhirnya menuju satu sebaran posterior tertentu. Perhitungan sebaran posterior menghasilkan sampel-sampel besaran posterior. Akhirnya, parameter dari sebaran posterior dapat diduga.

Prosedur MCMC yang terkenal adalah Gibbs bersyarat (*Gibbs Conditionals*). Menurut Rao (2003) bentuk Gibbs bersyarat untuk model Log-Normal adalah :

- i.  $[\hat{\mu} | \theta_i, \sigma^2, y_i] \sim N\left(\frac{1}{m} \sum_i \xi_i, \frac{\sigma^2}{m}\right)$

$$\text{ii. } [\hat{\sigma}^2 | \theta_i, \hat{\mu}, y_i] \sim G\left(\frac{m}{2} + a, \frac{1}{2} \sum_i (\xi_i - \hat{\mu})^2 + b\right)$$

$$\text{iii. } f(\theta_i | \hat{\mu}, \hat{\sigma}^2, y_i) \propto \theta_i^{y_i - 1} \exp\left[-\theta_i - \frac{1}{2\hat{\sigma}^2} (\xi_i - \hat{\mu})^2\right] \quad (4.3)$$

Pendugaan parameter  $\hat{\mu}$  dan  $\sigma^2$  dibangkitkan secara langsung dari (i) dan (ii) pada (4.3). Sementara itu, bagian (iii) persamaan (4.3) dinyatakan sebagai  $f(\theta_i | \hat{\mu}, \hat{\sigma}^2, y_i) \propto k(\theta_i) h(\theta_i | \hat{\mu}, \hat{\sigma}^2)$  dimana :

$$\text{i. } h(\theta_i | \hat{\mu}, \hat{\sigma}^2) \propto g'(\theta_i) \exp\left\{\frac{-(\xi_i - \hat{\mu})^2}{2\hat{\sigma}^2}\right\}$$

$$\text{dengan } g'(\theta_i) = \frac{\partial g(\theta_i)}{\partial \theta_i} \text{ dan}$$

$$g(\theta_i) = \log(\theta_i)$$

$$\text{ii. } k(\theta_i) = \exp(-e_i \theta_i) \theta_i^{y_i} \quad (4.4)$$

Sesuai dengan persamaan (4.3),  $\hat{\mu}$  dan  $\hat{\sigma}^2$  mengikuti distribusi yang standar yaitu *Multivariat Normal* dan *Gamma*. Sehingga nilai kedua parameter tersebut dapat ditaksir melalui pembangkitan sampel acak. Menurut Rao (2003), nilai rata-rata *Hierarchical Bayes* akan diduga melalui algoritma M-H (*gibbs sampling Metropolis-Hasting*), yaitu sebagai berikut :

1. Dibangkitkan  $\xi_i \sim iid N(\hat{\mu}, \hat{\sigma}^2)$  lalu dicari nilai  $\theta_i^{(0)} = g^{-1}(\xi_i)$

$$2. \text{ Dihitung nilai } \theta_i^* = \frac{1}{\theta_i^{(0)}} \exp\left\{\frac{-(\xi_i - \hat{\mu})^2}{2\hat{\sigma}^2}\right\}$$

3. Dihitung probabilitas penerimaan :

$$a(\theta_i^{(d)}, \theta_i^*) = \min\left\{\frac{k(\theta_i^*)}{k(\theta_i^{(d)})}, 1\right\};$$

$$d = 0, 1, \dots, D$$

4. Dibangkitkan  $u$  dari sebaran Seragam(0,1)

$$5. \text{ Dipilih } \theta_i^{(d+1)} = \theta_i^* \text{ jika } u \leq a(\theta_i^{(d)}, \theta_i^*)$$

6. Diulangi langkah 3 sampai 4, sampai diperoleh D sampel.

Setelah dilakukan simulasi M – H, maka diperoleh barisan penduga rata-rata sebagai berikut :

$$\{\theta_1^{(d)}, \dots, \theta_m^{(d)} ; d = 1, \dots, D\} \quad (4.5)$$

Kemudian besaran posterior yang diamati dapat dihitung. Penduga rata-rata *Hierarchical Bayes* ( $\hat{\theta}_i^{HB}$ ) adalah

$$\hat{\theta}_i^{HB} \approx \frac{1}{D} \sum_{k=1}^D \theta_i^{(d)} \quad (4.6)$$

(4.6)

Sedangkan varian penduga rata-rata *Hierarchical Bayes* ( $V(\theta_i^{HB} | \hat{\theta})$ ) adalah

$$V(\hat{\theta}_i^{HB} | \hat{\theta}) = \frac{1}{D-1} \sum_{d=1}^D (\theta_i^{(d)} - \hat{\theta}_i^{HB})^2 \quad (4.7)$$

(4.7)

## IV. TELADAN PENERAPAN

### 4.1 Data

Data pada penelitian ini adalah data yang dibangkitkan pada bahasa pemrograman R-3.2.2 dengan  $N = 36$ . Variabel  $y$  pada data penelitian merupakan data yang berdistribusi Poisson yang diperoleh dari banyaknya  $n$  sampel pada setiap area ke- $i$  yang masing-masing diperoleh menggunakan algoritma pada 4.2.1 no 1 dan 2.

### 4.2 Perhitungan Parameter Rata-rata ( $\theta_i$ )

Nilai parameter rata-rata ( $\theta_i$ ) yang diperoleh dari persamaan (2.6) menggunakan program *Microsoft Excel 2007* pada data penelitian ini ditunjukkan oleh tabel berikut :

*Excel 2007* pada data penelitian ini ditunjukkan oleh tabel berikut :

**Tabel 4.2.** Nilai parameter rata-rata ( $\theta_i$ )

Area ke-	$\theta_i$
1	0,900
2	1,331
3	1,234
4	1,525
5	1,258
6	1,039
7	1,484
8	1,093
9	0,915
10	1,093
11	1,438
12	0,726

Area ke-	$\theta_i$
13	0,924
14	0,959
15	0,928
16	0,702
17	1,076
18	0,876
19	1,273
20	1,062
21	1,273
22	0,911
23	0,932
24	1,107

Area ke-	$\theta_i$
25	0,932
26	0,548
27	1,302
28	0,899
29	0,373
30	0,935
31	0,971
32	0,751
33	0,388
34	0,702
35	1,011
36	1,120

### 4.3 Perhitungan Nilai Penduga Rata-Rata *Hierarchical Bayes* ( $\hat{\theta}_i^{HB}$ )

Setelah melakukan perhitungan parameter nilai rata-rata ( $\theta_i$ ), perhitungan nilai rata-rata *Hierarchical Bayes* ( $\hat{\theta}_i^{HB}$ ) dapat dilakukan diawali dengan menghitung nilai  $\mu$  dan dilanjutkan dengan menghitung nilai  $\sigma^2$ . Hasil perhitungan nilai  $\mu$  menggunakan bahasa pemrogram R-3.2.2 adalah **-0.05811381**. Hasil perhitungan nilai  $\sigma^2$  menggunakan bahasa pemrogram R-3.2.2 adalah **0.02064554**.

Berdasarkan nilai  $\hat{\mu}$  dan  $\hat{\sigma}^2$  yang diperoleh, nilai rata-rata *Hierarchical Bayes* ( $\hat{\theta}_i^{HB}$ ) diperoleh dengan menggunakan algoritma M-H dalam bahasa pemrograman R-3.2.2 ditunjukkan oleh tabel berikut :

**Tabel 4.4.** Nilai pendugaan rata-rata *Hierarchical Bayes* ( $\hat{\theta}_i^{HB}$ )

Area ke-	$\hat{\theta}_i^{HB}$	Area ke-	$\hat{\theta}_i^{HB}$	Area ke-	$\hat{\theta}_i^{HB}$
1	0,724	13	0,999	25	0,869
2	0,760	14	0,990	26	0,600
3	0,989	15	0,982	27	0,949
4	0,925	16	0,815	28	1,130
5	0,957	17	0,695	29	0,726
6	0,994	18	0,930	30	0,892
7	0,962	19	0,847	31	0,804
8	0,962	20	1,169	32	0,687
9	1,000	21	1,005	33	0,433
10	0,982	22	1,034	34	0,642
11	1,094	23	0,890	35	0,908
12	0,803	24	0,796	36	0,818

**Tabel 4.6.** Statistika deskriptif penduga langsung dan penduga metode *Hierarchical Bayes*

	Minimum	Median	Maksimum	Rata-rata
$(\theta_i)$	0,373	0,965	1,525	0,999
$(\hat{\theta}_i^{HB})$	0,251	0,951	1,825	0,894
$RMSE(\theta_i)$	0,264	0,421	0,539	0,423
$RMSE(\hat{\theta}_i^{HB})$	0	0,060	0,389	0,102

Pada Tabel 4.6 diperoleh informasi bahwa secara rata-rata penggunaan metode *Hierarchical Bayes* dengan fungsi prior Log-Normal memberikan ketelitian yang lebih baik yaitu dengan nilai  $RMSE$  yang lebih kecil dibanding pendugaan langsung.

#### IV. PENUTUP

##### 4.1 Kesimpulan

Berdasarkan hasil pembahasan mengenai pemodelan kejadian poisson pada *Small Area Estimation* menggunakan metode *Hierarchical Bayes* dengan fungsi prior Log-Normal, diperoleh kesimpulan sebagai berikut :

1. Langkah-langkah pemodelan kejadian poisson pada *Small Area Estimation* menggunakan metode *Hierarchical Bayes* dengan fungsi prior Log-Normal dimulai dari pendugaan nilai parameter rata-rata secara langsung, kemudian untuk perhitungan pendugaan rata-rata *Hierarchical Bayes* dibantu dengan teknik MCMC, yaitu :

- a. Menentukan nilai  $\hat{\mu} \sim N\left(\frac{1}{m} \sum_i \xi_i, \frac{\sigma^2}{m}\right)$
- b. Menentukan nilai  $\hat{\sigma}^2 \sim G\left(\frac{m}{2} + a, \frac{1}{2} \sum_i (\xi_i - \hat{\mu})^2 + b\right)$
- c. Setelah menentukan  $\hat{\mu}$  dan  $\hat{\sigma}^2$ , pendugaan parameter rata-rata *Hierarchical Bayes* ( $\hat{\theta}_i^{HB}$ ) dihitung menggunakan algoritma M-H, kemudian menghitung nilai rata-rata posteriornya yaitu :  $\hat{\theta}_i^{HB} \approx \frac{1}{D} \sum_{k=1}^D \theta_i^{(d)}$ .

d. Nilai *RMSE* digunakan sebagai ukuran keakuratan dan validasi pendugaan rata-rata, dengan menunjukkan nilai *RMSE* yang mendekati nol (0).

2. Pada penelitian ini diperoleh bahwa pendugaan parameter dengan menggunakan metode *Hierarchical Bayes* lebih baik dibanding dengan pendugaan secara langsung yang ditunjukkan dengan nilai *RMSE* yang lebih kecil.

#### 4.2 Saran

Penelitian serupa dapat dilanjutkan untuk kasus pemodelan kejadian poisson pada *Small Area Estimation* menggunakan metode *Hierarchical Bayes* menggunakan fungsi prior Log-Normal dengan peubah penyerta atau dengan melakukan pemodelan untuk area unit level.

#### DAFTAR PUSTAKA

- Bain, L. J. and Engelhardt, M., Introduction To Probability And Mathematical Statistics, Duxbury Press, United States of America, 1992
- Berger, O.J. 1985. Statistical Decision Theory and Bayesian Analysis. Springer, New York.
- Datta, G.S, Ghosh, M. 1991. Bayesian Prediction In Linear Models: Application To Small Area Estimation. Ann. Statist. 19, 1746-1770.
- Ghosh, M, dan J.K.N. Rao. 1994. *Small Area Estimation: An Appraisal. Statistical Science* 9: 55-93.
- Hadi, A. F, dan Nusyirwan. 2008. *Penduga Maksimum Likelihood Untuk Parameter Dispersi Model Poisson-Gamma Dalam Konteks Penduga Area Kecil*. Penelitian ini disampaikan pada Seminar Nasional Matematika dan Pendidikan Matematika. FMIPA IPB.
- Hajarisman, N. 2013. *Pemodelan Area Kecil Untuk Menduga Angka Kematian Bayi Melalui Pendekatan Model Regresi Poisson Bayes Berhierarchy Dua-Level*. [Disertasi]. Bogor: Departemen Statistika, FMIPA Institut Pertanian Bogor.
- Handayani, D. 2015. *Pendugaan Parameter Pada Suatu Small Area dengan Memperhatikan Masalah Ketergantungan Spasial*. Bogor: IPB.
- Hastings. W.K. 1970. *Monte Carlo Sampling Methods Using Marcov Chain And Their Applications*. Biometrika 57. 1317-1340.
- Kismiantini. 2007. *Pendekatan Bayes Empirik Pada Pendugaan Statistik Area Kecil Berbasis Model Poisson-Gamma Dengan Peubah Penyerta*. Makalah ini disampaikan pada Seminar Nasional Penelitian, Pendidikan dan Penerapan MIPA, FMIPA Universitas Negeri Yogyakarta. [25 Agustus 2007]
- Kismiantini. 2008. *Pendugaan Berbasis Model Untuk Kasus Biner Pada Small Area Estimation*. Makalah ini disampaikan pada Seminar Nasional Penelitian, Pendidikan dan Penerapan MIPA, FMIPA Universitas Negeri Yogyakarta. [28 November 2008]
- Kismiantini. 2010. *Penerapan Metode Bayes Empirik Pada Pendugaan Area Kecil Untuk Kasus Biner*. Makalah ini disampaikan pada Seminar Nasional Penelitian, Pendidikan dan Penerapan MIPA, FMIPA Universitas Negeri Yogyakarta. [15 Mei 2010]
- Kurnia, A. 2009. *Prediksi Terbaik Empirik untuk Model Transformasi Logaritma di Dalam Pendugaan Area Kecil dengan Penerapan Pada Data Susenas*. [Disertasi]. Bogor: FMIPA IPB.
- Maiti, T. 1998. *Hierarchical Bayes Estimation of Mortality Rates For Disease Mapping*. Journal of statistical Planning and Inference. 69. 339-349.

- Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E. 1953. *Equation of State Calculations by Fast Computing Machines*. J.Chem. Phys. 21, 1087-1092.
- Nadhiroh, I. M. 2008. *Zero Inflated Negative Binomial Models In Small Area Estimation*. Penelitian ini disampaikan pada Seminar Nasional Matematika dan Pendidikan Matematika. FMIPA IPB.
- Notoatmodjo, S. 2005. *Metodologi Penelitian Kesehatan*. Rineka Ilmu. Jakarta
- Noviyanti, R.A. 2014. *Pendekatan Small Area Estimation Pada Scan Statistic Untuk Pendeteksian Kantong Kemiskinan*. Penelitian ini disampaikan pada Prosiding Seminar Nasional Matematika, Universitas Jember. [19 November 2014].
- Rao, J.N.K. 2003. *Small Area Estimation*. New Jersey: John Wiley & Sons, Inc.
- Rumiati, A.T. 2012. *Empirical Bayesian Method For Teh Estimation Of Literacy Rate At Sub-District Level Case Study: Sumenep District Of East Java Province*. IPTEK, *The journal for Technology and Science*, Vol 23, Number 1, February 2012.
- Sunandi, E. 2013. *Model Spasial Bayes dalam Pendugaan Area Kecil dengan Peubah Respon Biner*. Kumpulan Makalah Seminar Semirata 2013. FMIPA Lampung, p: 179-183.
- Walpole, E. R.1995. *Pengantar Statistika*. Gramedia Pustaka Utama: Jakarta.

# PENARIKAN SAMPEL BERDASARKAN PERINGKAT (*RANKED SET SAMPLING*) PADA ANALISIS REGRESI LINIER BERGANDA

Dewanty Manurung<sup>1</sup>, Sigit Nugroho<sup>2</sup>, dan Fachri Faisal<sup>2</sup>

<sup>1</sup>Mahasiswa Jurusan Matematika Fakultas MIPA Unuversitas Bengkulu

<sup>2</sup>Dosen Pembimbing Jurusan Matematika Fakultas MIPA Universitas Bengkulu  
[dewanty\\_manurung@yahoo.com](mailto:dewanty_manurung@yahoo.com)

## ABSTRAK

Penarikan sampel pada suatu penelitian bertujuan untuk mengurangi waktu dan biaya, sehingga jumlah sampel yang kecil diharapkan mampu mewakili populasi yang ada. *Simple Random Sampling* (SRS) merupakan metode penarikan sampel yang sering dilakukan oleh masyarakat pada umumnya. Namun, SRS tidak menjamin bahwa sampel yang diperoleh telah mewakili data populasi karena tingkat presisinya tidak cukup tinggi. *Ranked Set Sampling* merupakan suatu metode penarikan sampel berdasarkan peringkat yang mempertimbangkan variabel konkomitan pada data yang ingin diteliti. Nilai *Relative Precision* digunakan sebagai acuan pemilihan metode terbaik. Tujuan dari penelitian ini adalah untuk mempelajari RSS pada regresi linier berganda, untuk mengetahui  $Var(\hat{\beta}_i), i = 0,1,2$ , dan untuk membandingkan SRS dengan RSS. Hasil penelitian menunjukkan bahwa RSS pada regresi linier berganda yang dilakukan dengan mengurutkan data berdasarkan variabel  $X_1$  kemudian berdasarkan variabel  $X_2$  digunakan untuk mengestimasi  $Var(\hat{\beta}_0)$ . Nilai  $Var(\hat{\beta}_i), i = 0,1,2$  mengalami penurunan untuk  $m$  yang sama dan  $n$  yang semakin besar. Nilai  $RP(\hat{\beta}_0)$  untuk  $n = 30, 60, 90$  lebih besar dari 1, sehingga dapat disimpulkan bahwa RSS lebih efisien dari pada SRS.

**Kata kunci:** SRS, RSS, *Relative Precision*.

## ABSTRACT

Sampling on a research aims to reduce the time and the costs, so that the small sample size is expected to represent the population. *Simple Random Sampling* (SRS) is a sampling method that is often done by public. However, SRS does not guarantee that the sample obtained was representative of the population because of the level of data precision is not high enough. *Ranked Set Sampling* is a sampling method based on the ranking variable concomitant consider the data to be examined. *Relative Precision* value is used as a reference for the selection of the best method. The aims of this research is to study the RSS in multiple linear regression, to find  $Var(\hat{\beta}_i), i = 0,1,2$  and to compare SRS with RSS method. The results showed that the RSS in multiple linear regression were performed to sort the data based on a variable by variable concomitant.  $Var(\hat{\beta}_i), i = 0,1,2$  decreased for the same  $m$  and  $n$  is greater.  $RP(\hat{\beta}_0), i = 0,1,2$  greater from 1, so it can be conclude that RSS is more efficient than SRS.

**Keywords:** SRS, RSS, *Relative Precision*

## 1. PENDAHULUAN

Dalam suatu penelitian, untuk menduga nilai parameter populasi diperlukan penarikan sampel atau contoh yang bertujuan untuk mengurangi waktu dan biaya pengumpulan data. Penarikan sampel harus dilakukan secara hati-hati dan dengan pertimbangan yang tepat.

*Simple Random Sampling (SRS)* merupakan salah satu tipe *sampling probability* yang dilakukan secara acak (*random*) sehingga setiap kasus atau elemen dalam populasi memiliki kesempatan yang sama besar untuk dipilih sebagai sampel penelitian. SRS dapat digunakan apabila keadaan variabel yang akan diteliti relatif homogen dan tersebar merata di seluruh populasi. Namun, pemilihan secara acak tidak menjamin bahwa sampel yang diperoleh telah mewakili data populasi karena tingkat presisinya tidak cukup tinggi (Sari, 1993). Pada tahun 1952, McIntyre memperkenalkan sebuah teknik penarikan sampel berdasarkan peringkat atau *Ranked Set Sampling (RSS)*. Nilai rata-rata populasi pengamatan yang diukur berdasarkan metode RSS umumnya lebih representatif daripada yang diperoleh berdasarkan *simple random sampling* (Sorka, 2008).

Pada regresi linier berganda, selain menggunakan data variabel respon juga menggunakan data tambahan (variabel konkomitan) lebih dari satu variabel bebas di mana, antara variabel respon dan variabel bebas harus memiliki korelasi (hubungan). Pada penelitian ini, penulis tertarik untuk membahas penarikan sampel berdasarkan peringkat atau *ranking set sampling* pada analisis regresi linier berganda.

## 2. RANKED SET SAMPLING (RSS)

Langkah-langkah penarikan sampel:

- a. mengambil  $m$  elemen dari  $N$  populasi dengan menggunakan metode *Simple Random Sampling Without Replacement (SRSWOR)*. Menurut Cochran (1991), SRSWOR merupakan teknik penarikan sampel di mana sebuah bilangan yang telah diambil tidak diikutsertakan lagi dalam populasi untuk seluruh penarikan berikutnya, dan menurut Patil (2002) penarikan  $m$  sebaiknya menggunakan ukuran kecil misalnya 3 atau 4 untuk meminimalkan *error* dalam pemberian peringkat. Penarikan sampel pada langkah pertama diulang

secara bebas sebanyak  $m$  kali yang menghasilkan  $m$  set (kelompok) sampel saling bebas yang berukuran  $m$ .

- b. Memberikan peringkat pada masing-masing sampel dalam setiap kelompok.
- c. Ambil elemen dengan peringkat terkecil dari kelompok pertama, elemen dengan peringkat kedua terkecil dari kelompok yang kedua, dan seterusnya dipilih sampai elemen dengan peringkat terbesar dari kelompok ke- $m$ , sehingga diperoleh sampel sebanyak  $m$ .
- d. Ulang seluruh proses di atas sebanyak  $r$  kali sampai  $n$  sampel yang dibutuhkan terpenuhi di mana  $n = mr$ .

Misalkan  $x_k$  merupakan notasi dari nilai variabel  $x$  untuk elemen populasi ke- $k$ ,  $k = 1, 2, \dots, N$  dan  $x_{ij}$  untuk  $i = 1, 2, \dots, m; j = 1, 2, \dots, r$  merupakan notasi dari nilai variabel  $x$  yang diambil sebagai sampel ke- $i$  pada pengulangan ke- $j$ . Rata-rata RSS ketika semua prosedur diulang  $r$  kali, maka nilai rata-ratanya adalah

$$\bar{x}_{[m]r} = \frac{1}{mr} \sum_{j=1}^r \sum_{i=1}^m x_{i(i:m)j} \quad (1)$$

di mana  $x_{i(i:m)j}$  merupakan urutan statistik ke- $i$  dari sampel ke- $i$  pada ulangan ke- $j$ . Nilai rata-rata  $\bar{x}_{[m]r}$  pada RSS dari populasi berhingga dengan  $r$  kali pengulangan merupakan estimator takbias dari rata-rata populasi ( $\mu$ ). Nilai estimasi varians populasi menggunakan metode RSS adalah

$$s^2 = \frac{1}{mr - 1} \sum_{i=1}^m \sum_{j=1}^r (x_{i:j} - \mu_{SRS})^2 \quad (2)$$

di mana  $\mu_{SRS} = \frac{1}{mr} \sum_{i=1}^m \sum_{j=1}^r x_{i:j}$ , sehingga

$$E(s_{RSS}^2) = \sigma^2 + \frac{1}{m(mr - 1)} \sum_{i=1}^m (\mu_{[i]} - \mu)^2 \quad (3)$$

dan nilai varians rata-rata RSS dari populasi berhingga merupakan estimator takbias dari varians populasi.

$$Var(\bar{x}_{RSS}) = \frac{1}{n} \left\{ \left(1 - \frac{1}{N}\right) \sigma^2 - \frac{1}{n} \sum_{i=1}^n (E(x_{i:n}) - \mu)^2 \right\} \quad (4)$$

### 3. RSS KONKOMITAN

Langkah-langkah penarikan sampel pada RSS dengan satu variabel konkomitan adalah mengambil  $m$  sample dari  $N$  populasi variabel  $X$  menggunakan teknik SRSWOR sebanyak  $m$  kali, kemudian elemen  $X$  diberikan peringkat pada masing-masing kelompok. Kelompok sampel pertama yang berukuran  $m$  elemen, nilai  $Y$  yang berpasangan dengan  $X$  yang berperingkat paling kecil dipilih, sampel kedua yang memiliki  $m$  elemen, nilai  $Y$  yang berpasangan dengan  $X$  yang berperingkat terkecil kedua dipilih, dan seterusnya dilakukan sampai  $Y$  yang berpasangan dengan  $X$  berperingkat tertinggi dari sampel ke- $m$  dipilih untuk dilakukan perhitungan. Keseluruhan proses diatas dapat diulang sebanyak  $r$  kali sampai  $n$  sampel yang dibutuhkan terpenuhi di mana  $n = mr$ .

### 4. RSS PADA REGRESI LINIER BERGANDA DUA VARIABEL BEBAS

Langkah-langkahnya adalah sebagai berikut (Chen dkk, 2004) :

1. Misalkan  $k, m$  merupakan bilangan bulat positif.
2. Ambil  $m$  kelompok-kelompok saling bebas, di mana setiap kelompok berukuran  $k$  diambil dari populasi.
3. Urutkan elemen-elemen dalam setiap kelompok berdasarkan variabel  $X_1$ .
4. Bentuk kelompok  $m$  pertama yang baru, di mana elemen-elemen yang ada di dalamnya merupakan elemen-elemen yang memiliki peringkat pertama yang diurutkan berdasarkan variabel  $X_1$  pada setiap kelompok yang ada pada langkah 2. Bentuk kelompok  $m$  kedua yang baru, di mana elemen-elemen yang ada di dalamnya merupakan elemen-elemen yang memiliki peringkat kedua yang diurutkan berdasarkan variabel  $X_1$  pada setiap kelompok yang ada pada langkah 2, dan seterusnya .
5. Prosedur RSS berdasarkan variabel pertama selesai. Penarikan sampel berdasarkan variabel  $X_2$  adalah dengan menggunakan elemen-elemen yang telah diperoleh pada peringkatan berdasarkan variabel  $X_1$ . Elemen-elemen yang telah ada pada langkah 4 diurutkan berdasarkan variabel  $X_2$  untuk setiap kelompok.
6. Pilih elemen yang mempunyai peringkat pertama pada kelompok pertama yang telah diurutkan berdasarkan variabel  $X_2$ . Pilih elemen yang mempunyai

peringkat kedua pada kelompok kedua yang telah diurutkan berdasarkan variabel  $X_2$ , demikian seterusnya hingga kelompok ke- $k$ . Variabel  $Y$  dari elemen-elemen yang terpilih diambil untuk dilakukan perhitungan.

7. Ulangi langkah 1 sampai 6 sebanyak  $r$  kali ulangan, sehingga diperoleh data  $n = mr$ .

Model regresi linier berganda dua variabel bebas untuk penarikan sampel berperingkat dengan ukuran sampel  $n = mr$  adalah

$$\mathbf{Y}_R = \mathbf{X}_R \boldsymbol{\beta}_R + \boldsymbol{\varepsilon}_R \quad (5)$$

di mana

$$\mathbf{Y}_R = \begin{bmatrix} Y_{[1]1} \\ Y_{[2]1} \\ \vdots \\ Y_{[m]1} \\ \vdots \\ Y_{[1]r} \\ Y_{[2]r} \\ \vdots \\ Y_{[m]r} \end{bmatrix}, \quad \mathbf{X}_R = \begin{bmatrix} 1 & X_{1[1]1} & X_{2[1]1} \\ 1 & X_{1[2]1} & X_{2[2]1} \\ \vdots & \vdots & \vdots \\ 1 & X_{1[m]1} & X_{2[m]1} \\ \vdots & \vdots & \vdots \\ 1 & X_{1[1]r} & X_{2[1]r} \\ 1 & X_{1[2]r} & X_{2[2]r} \\ \vdots & \vdots & \vdots \\ 1 & X_{1[m]r} & X_{2[m]r} \end{bmatrix}, \quad \boldsymbol{\beta}_R = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}, \quad \boldsymbol{\varepsilon}_R = \begin{bmatrix} \varepsilon_{[1]1} \\ \varepsilon_{[2]1} \\ \vdots \\ \varepsilon_{[m]1} \\ \vdots \\ \varepsilon_{[1]r} \\ \varepsilon_{[2]r} \\ \vdots \\ \varepsilon_{[m]r} \end{bmatrix}$$

Keterangan:

- $\mathbf{X}_R$  merupakan matriks variabel bebas berdimensi  $mr \times 3$  dengan elemennya adalah  $(1, X_{1[i]j}, X_{2[i]j})$ ,  $i = 1, 2, \dots, m; j = 1, 2, \dots, r$ .
- $X_{p[i]j}$  merupakan notasi untuk urutan statistik ke- $i$  pada variabel  $X_p$  kelompok ke- $i$  dan ulangan ke- $j$ .
- $\mathbf{Y}_R$  merupakan vektor variabel terikat berdimensi  $mr \times 1$  dengan elemennya adalah  $Y_{[i]j}$ .
- $Y_{[i]j}$  merupakan nilai dari variabel terikat untuk variabel bebas ke- $i$  pada kelompok ke- $i$  dan ulangan ke- $j$ .
- $\boldsymbol{\beta}_R$  merupakan vektor parameter berdimensi  $3 \times 1$ .
- $\boldsymbol{\varepsilon}_R$  merupakan vektor acak *error* berdimensi  $nr \times 1$  dengan  $E(\boldsymbol{\varepsilon}_R) = \mathbf{0}$ ,  $Var(\boldsymbol{\varepsilon}_R) = \sigma^2 \mathbf{I}$  dan  $Cov(\boldsymbol{\varepsilon}_R, \mathbf{X}_R) = \mathbf{0}$ .

Estimasi dari parameter  $\beta$  berdasarkan RSS adalah dengan menggunakan model berikut

$$\widehat{\beta}_{RSS} = (\mathbf{X}'_R \mathbf{X}_R)^{-1} \mathbf{X}'_R \mathbf{Y}_R \quad (6)$$

Varians dari estimatornya adalah

$$\text{Var}(\widehat{\beta}_{RSS}) = \sigma^2 (\mathbf{X}'_R \mathbf{X}_R)^{-1} \quad (7)$$

di mana  $\sigma^2$  dinyatakan dengan persamaan berikut:

$$\sigma_{RSS}^2 = \frac{\mathbf{Y}'_R \mathbf{Y}_R - \widehat{\beta}'_R \mathbf{X}'_R \mathbf{Y}_R}{mr - (p + 1)} \quad (8)$$

(Ozdemir & Esin, 2007)

Menurut Supranto (2007) dalam Wijayanti (2013), *Relative Precision* dari pengambilan sampel pada regresi linier berganda dengan dua variabel bebas adalah

$$RP(\widehat{\beta}_{(l-1)}) = \frac{\text{Var}(\widehat{\beta}_{SRS})_l}{\text{Var}(\widehat{\beta}_{RSS})_l} = \frac{\sigma^2 [(\mathbf{X}'_S \mathbf{Y}_S)^{-1}]_l}{\sigma^2 [(\mathbf{X}'_R \mathbf{Y}_R)^{-1}]_l}, \quad l = 1, 2, 3 \quad (9)$$

Apabila nilai  $\text{Var}(\widehat{\beta}_{RSS})_l \leq \text{Var}(\widehat{\beta}_{SRS})_l$  untuk  $l = 1, 2, 3$  dan nilai  $RP(\widehat{\beta}_{(l-1)}) \geq 1$  maka nilai penduga parameter RSS ( $\widehat{\beta}_{RSS}$ ) lebih efisien atau minimal sama efisiennya dari nilai penduga parameter SRS ( $\widehat{\beta}_{SRS}$ ) dengan ukuran sampel yang sama yaitu  $n = mr$ .

## 5. SRS PADA REGRESI LINIER BERGANDA DUA VARIABEL BEBAS

*Simple Random Sampling* (SRS) atau penarikan sampel acak sederhana adalah sebuah metode untuk memilih  $n$  unit dari  $N$  elemen populasi sehingga setiap elemen dari  ${}_N C_n$  sampel yang berbeda mempunyai kesempatan yang sama untuk dipilih (Cochran, 1991). Berdasarkan penarikan sampel acak sederhana dengan ukuran  $n$  dari data berdistribusi normal multivariat pada regresi linier berganda dengan dua variabel bebas ( $Y, X_1, X_2$ ) dalam notasi matriks dapat ditulis sebagai berikut

$$\mathbf{Y}_s = \mathbf{X}_s \beta_s + \varepsilon_s \quad (10)$$

di mana

$$Y_S = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \quad X_S = \begin{bmatrix} 1 & X_{11} & X_{12} \\ 1 & X_{21} & X_{22} \\ \vdots & \vdots & \vdots \\ 1 & X_{n1} & X_{n2} \end{bmatrix}, \quad \beta_S = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}, \quad \varepsilon_S = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Keterangan:

- $Y_S$  merupakan vektor variabel terikat berdimensi  $n \times 1$ .
- $X_S$  merupakan matriks variabel bebas berdimensi  $n \times 3$  pada SRS.
- $\beta_S$  merupakan vektor parameter berdimensi  $3 \times 1$ .
- $\varepsilon_S$  merupakan vektor acak *error* berdimensi  $n \times 1$  dengan  $E(\varepsilon_S) = \mathbf{0}$ ,  $Var(\varepsilon_S) = \sigma^2 I$  dan  $Cov(\varepsilon_S, X_S) = \mathbf{0}$ .

Estimasi dari parameter  $\beta$  berdasarkan SRS adalah dengan menggunakan model berikut

$$\hat{\beta}_{SRS} = (X_S' X_S)^{-1} X_S' Y_S \quad (11)$$

Varians dari estimatornya adalah

$$Var(\hat{\beta}_{SRS}) = \sigma^2 (X_S' X_S)^{-1} \quad (12)$$

Estimator takbias dari  $\sigma^2$  dinyatakan sebagai berikut

$$\sigma_{SRS}^2 = \frac{Y_S' Y_S - \hat{\beta}_S' X_S' Y_S}{n - (p + 1)} \quad (13)$$

(Ozdemir & Esin, 2007)

## 6. DESKRIPSI DATA

Data yang digunakan dalam penelitian ini merupakan data yang dibangkitkan dengan menggunakan bantuan program *Microsoft Excel*. Data tersebut merupakan data regresi linier beganda dengan dua variabel bebas yaitu  $X_1$  dan  $X_2$ , di mana jumlah datanya dianggap sebagai populasi dengan  $N = 130$ . Nilai parameter  $\beta$  dari data regresi hasil bangkitannya adalah  $\beta_0 = -90.1116$ ,  $\beta_1 = 11.1905$ , dan  $\beta_2 = 15.8418$ . Sedangkan nilai  $Var(\beta_0) = 3200.1107$ ,  $Var(\beta_1) = 0.3017$ , dan  $Var(\beta_2) = 3.4335$ . Penarikan sampel dilakukan dengan mengambil  $n = 30, 60$ , dan  $90$ .

## 7. HASIL DAN PEMBAHASAN

Nilai  $Var(\hat{\beta}_i)$ ,  $i = 0, 1, 2$  pada penarikan sampel dengan SRS dengan pengulangan 100 kali yaitu:

**Tabel 1.**  $Var(\hat{\beta}_i)$  pada Metode SRS

$n$	$Var(\beta_0)$	$Var(\beta_1)$	$Var(\hat{\beta}_2)$
30	13964.76	1.36687	15.3424
60	6961.468	0.648931	7.495835
90	4610.899	0.431987	4.980041

Berdasarkan Tabel 3 diperoleh bahwa  $Var(\hat{\beta}_i), i = 0,1,2$  dengan metode SRS mengalami penurunan untuk nilai  $n$  yang semakin besar. Hal ini disebabkan karena semakin besarnya jumlah sampel yang digunakan maka variansi dari data akan semakin kecil.

Penarikan Sampel dengan Metode RSS pada Regresi Linier Berganda penulis mengasumsikan  $k = m$ . Untuk  $n = 30$  diambil  $m = 3, r = 10$ , sebagai pembandingnya penulis juga mengambil  $m = 5, r = 6$  dan  $m = 6, r = 5$ . Untuk jumlah sampel  $n = 60$ , diambil  $m = 3, r = 20$ ,  $m = 5, r = 12$  dan  $m = 6, r = 10$ . Demikian juga untuk jumlah sampel  $n = 90$ , diambil  $m = 3, r = 30$ ,  $m = 5, r = 18$  dan  $m = 6, r = 15$ . Nilai  $Var(\hat{\beta}_i)$  untuk masing-masing  $n$  dapat dilihat pada Tabel 2.

**Tabel 2.**  $Var(\hat{\beta}_i)$  pada Metode RSS

$n$	$Var(\hat{\beta}_i)$	$m = 3$	$m = 5$	$m = 6$
30	$Var(\hat{\beta}_0)$	12674.42	12393.39	13294.66
	$Var(\hat{\beta}_1)$	1.836479	2.589608	2.633059
	$Var(\hat{\beta}_2)$	18.66993	24.14843	25.5343
60	$Var(\hat{\beta}_0)$	6397.612	5933.559	6694.454
	$Var(\hat{\beta}_1)$	0.884757	1.18610	1.1694
	$Var(\hat{\beta}_2)$	9.14460	11.31029	11.42837
90	$Var(\hat{\beta}_0)$	4197.361	4109.727	4403.244
	$Var(\hat{\beta}_1)$	0.571337	0.788122	0.787547
	$Var(\hat{\beta}_2)$	5.92314	7.537939	7.703791

Berdasarkan Tabel 2 diperoleh bahwa nilai  $Var(\hat{\beta}_i), i = 0,1,2$  mengalami penurunan pada  $n$  yang semakin meningkat. Semakin besar jumlah sampel yang digunakan, maka semakin kecil nilai  $Var(\hat{\beta}_i), i = 0,1,2$ . Namun, pada nilai  $n$  yang sama, nilai  $Var(\hat{\beta}_0)$  hanya mengalami penurunan pada  $m = 5$  sedangkan pada  $m = 6$

nilai  $Var(\hat{\beta}_0)$  selalu naik. Hal ini berarti  $m = 6$  memiliki *error* yang lebih tinggi pada pemberian peringkat saat data diurutkan berdasarkan variabel  $X_1$  dan variabel  $X_2$ .

### 8. Nilai *Relative Precision* (RP)

Perbandingan nilai  $Var(\hat{\beta}_i)$  pada SRS dengan nilai  $Var(\hat{\beta}_i)$  pada RSS menunjukkan ukuran yang digunakan untuk melihat metode terbaik dari kedua metode yang dilakukan pada masing-masing  $n$  dapat dilihat pada Tabel 3.

**Tabel 3.** Nilai  $RP(\hat{\beta}_i)$  RSS dan SRS

$n$	$RP(\hat{\beta}_i)$	m=3	m=5	m=6
30	$RP(\hat{\beta}_0)$	1.101807	1.12679	1.050404
	$RP(\hat{\beta}_1)$	0.744288	0.527829	0.519119
	$RP(\hat{\beta}_2)$	0.82177	0.635337	0.600854
60	$RP(\hat{\beta}_0)$	1.088135	1.173236	1.039886
	$RP(\hat{\beta}_1)$	0.733457	0.54708	0.554904
	$RP(\hat{\beta}_2)$	0.819701	0.662745	0.655897
90	$RP(\hat{\beta}_0)$	1.098523	1.121948	1.04716
	$RP(\hat{\beta}_1)$	0.756098	0.548122	0.548522
	$RP(\hat{\beta}_2)$	0.840777	0.660663	0.64644

Dari hasil yang diperoleh pada Tabel 3 terlihat bahwa untuk setiap  $n = 30, 60, 90$  pada saat  $m = 3$  ke  $m = 5$  nilai  $RP(\hat{\beta}_0)$  mengalami kenaikan, tetapi pada  $m = 6$  mengalami penurunan nilai. Hal ini dipengaruhi oleh penggunaan  $m$  yang besar akan memberikan *error* pada saat perangkaan tinggi, sehingga nilai  $RP(\hat{\beta}_0)$  yang dihasilkan semakin kecil. Menurut Ozdemir dan Esin (2007), penggunaan perangkaan terhadap  $X_1$  saja digunakan untuk mengestimasi nilai penduga paramer terhadap  $\hat{\beta}_0$  dan  $\hat{\beta}_1$ , sedangkan penggunaan perangkaan terhadap  $X_2$  saja digunakan untuk mengestimasi nilai penduga paramer terhadap  $\hat{\beta}_0$  dan  $\hat{\beta}_2$ . Pada penelitian ini penggunaan perangkaan terhadap  $X_1$  dan  $X_2$  menggunakan estimasi nilai penduga paramer terhadap  $\hat{\beta}_0$ . Jadi nilai  $RP(\hat{\beta}_i)$  yang dilihat adalah nilai  $RP(\hat{\beta}_0)$  saja. Artinya, berdasarkan nilai  $RP(\hat{\beta}_0)$  untuk  $n = 30$  dan  $m = 3, 5, 6$  penarikan sampel berdasarkan peringkat (RSS) lebih baik dari pada penarikan sampel acak biasa (SRS). Karena nilai  $RP(\hat{\beta}_0)$  untuk  $n = 30, 60, 90$  lebih besar dari 1 maka penarikan sampel dengan menggunakan metode RSS lebih baik dari pada metode SRS.

## 9. KESIMPULAN DAN SARAN

Berdasarkan hasil dan pembahasan yang telah dilakukan, dapat disimpulkan bahwa perangkingan terhadap variabel  $X_1$  dan  $X_2$  menghasilkan pendugaan parameter terhadap  $\hat{\beta}_0$  saja, nilai  $Var(\hat{\beta}_0)$  pada RSS dengan  $n$  yang sama mengalami penurunan dari  $m = 3$  ke  $m = 5$ , sedangkan nilai  $Var(\hat{\beta}_1)$  dan  $Var(\hat{\beta}_2)$  cenderung mengalami peningkatan nilai untuk  $n$  yang sama dan  $m$  yang semakin besar, dan nilai  $RP(\hat{\beta}_0)$  memiliki nilai yang meningkat pada data dengan  $m = 3$  dan  $m = 5$  sedangkan saat  $m = 6$  mengalami penurunan.

Untuk memperoleh sampel yang representatif dengan menggunakan metode RSS pada data yang memiliki jumlah populasi  $N = 130$  sebaiknya menggunakan  $n$  yang besar sehingga diperoleh  $Var(\hat{\beta}_i), i = 0,1,2$  yang lebih baik. Pembaca dapat melanjutkan tulisan ini dengan membahas tentang seberapa besar nilai  $m$  yang baik digunakan pada RSS sehingga sampel yang digunakan tetap representatif untuk setiap penarikan  $n$  sampel pada populasi berukuran  $N$ , dan sebaiknya menggunakan data populasi yang lebih besar.

## 10. DAFTAR PUSTAKA

- Chen, Z., Zhidong, B., & Bimal, K.S. 2004. *Ranked Set Sampling: Theory and Applications*. Springer-Verlag. New York, Inc.
- Cochran, W.G. 1991. *Teknik Penarikan Sampel*. Buku Terjemahan. Jakarta: UI Press.
- Ozdemir, Y.A. & Esin, A.A. 2007. Parameter Estimation in Multiple Linier Regression Models Using Ranked Set Sampling. *Commun.Fac.Sci.Univ.Ank.Series A1*. Vol. 56, no. 1, 7-20
- Sari, E.S. 1993. *Audience Research: Pengantar Studi Penelitian Terhadap Pembaca, Pendengar dan Pemirsa*. Yogyakarta: Andi Offset.
- Sorka, C. 2008. *Extending Ranked Set Sampling to Survey Methodology*. Columbus: The Ohio State University.
- Wijayanti, P.S. 2013. Pengambilan Sampel Berdasarkan Peringkat pada Analisis Regresi Linier Sederhana. *Jurnal Gaussian*. Vol. 2, no. 3, 209-218.

# PENENTUAN AWAL KEANGGOTAAN ANALISIS KLASTER NON HIRARKI (*K-MEANS*)

Monika Afriana<sup>1</sup>, Sigit Nugroho<sup>2</sup> dan Fachri Faisal<sup>2</sup>

<sup>1</sup>Mahasiswa Jurusan Statistika, Prodi Matematika, FMIPA, UNIB, Bengkulu

<sup>2</sup>Dosen Pembimbing Jurusan Statistika, Prodi Matematika, FMIPA, UNIB, Bengkulu  
[monikaafriana027@gmail.com](mailto:monikaafriana027@gmail.com)

## ABSTRAK

Penelitian ini bertujuan untuk menentukan awal keanggotaan analisis klaster non hirarki *K-Means* dan memberikan contoh penerapan penentuan awal keanggotaan analisis klaster non hirarki *K-Means*. Metode penelitian yang digunakan adalah studi literatur dan teladan penerapan yang diambil dari buku berjudul “Applied Multivariate Statistical Second Edition” karangan Johnson dan Winchern (2002: 741). Hasil penelitian menunjukkan bahwa dalam menentukan awal keanggotaan (jumlah klaster  $K$ ) pada klaster *K-Means* secara signifikan nilai  $K$  harus lebih kecil dari jumlah objek dalam himpunan data. Namun, perlu diperhatikan bahwa tidak terdapat aturan khusus dalam menentukan jumlah klaster  $K$ , terkadang jumlah klaster yang diinginkan tergantung pada subjektif seseorang. Algoritma *K-Means* tidak selalu menghasilkan hasil pengklasteran yang sama, salah satu penyebabnya karena ada unsur pengacakan pada titik pusat klaster .

**Kata kunci:** *K-Means*, Algoritma *K-Means*, Penentuan nilai  $K$ .

## 1. PENDAHULUAN

Analisis klaster merupakan salah satu teknik dalam analisis multivariat yang mempunyai tujuan utama untuk mengelompokkan objek-objek pengamatan menjadi beberapa kelompok berdasarkan karakteristik yang dimilikinya. (Nugroho, 2008). Proses pengelompokan data pada analisis klaster dapat dilakukan dengan dua metode yaitu metode hirarki dan metode non hirarki. Metode hirarki memulai pengelompokkan dengan dua atau lebih objek yang memiliki kesamaan paling dekat. Kemudian proses diteruskan ke objek lain yang mempunyai kedekatan kedua. Demikian seterusnya sehingga klaster akan membentuk semacam ‘pohon’ dimana ada hirarki (tingkatan) yang jelas antar objek, dari yang paling mirip sampai paling tidak mirip. Beberapa metode untuk proses klastering secara hirarki yaitu *single linkage*, *complete linkage*, *average linkage*, *ward’s method*, dan *centroid method* (Santoso, 2010).

Berbeda dengan metode hirarki, analisis klaster dengan metode non hirarki merupakan metode klaster yang menentukan jumlah klaster secara manual. Kekurangan metode ini adalah tidak disediakan gambar dendogram bentuk dari klaster untuk masing-masing sampel (Baroroh, 2012). Metode ini dimulai dengan menentukan terlebih dahulu jumlah klaster yang diinginkan. Setelah jumlah klaster diketahui, baru proses klaster dilakukan tanpa mengikuti proses hirarki. Metode ini disebut dengan Klaster *K-Means* (Santoso, 2010).

Pemecahan permasalahan pengklasteran dengan menggunakan metode *K-Means* ini secara umum dijalankan dengan menggunakan algoritma yang dikenal dengan algoritma *K-Means*. Algoritma *K-Means* adalah metode klastering berbasis jarak yang

membagi data ke dalam sejumlah klaster dan hanya bekerja pada atribut numerik. Algoritma ini pada awalnya mengambil sebagian dari banyaknya komponen dari populasi yang akan dijadikan sebagai pusat awal klaster (*centroid*). Pada langkah ini *centroid* dipilih secara acak dari sekumpulan populasi data (Martiana, dkk, 2011). Selanjutnya menghitung ukuran jarak dari masing-masing objek ke *centroid*. Setelah objek masuk pada *centroid* terdekat dan membentuk klaster baru, *centroid* baru ditentukan kembali dengan menghitung rata-rata objek pada *centroid* yang sama. Jika masih ada perbedaan dengan *centroid* yang sudah dibentuk, maka dilakukan perhitungan kembali *centroid* baru (Lathifaturrahmah, 2010).

Beberapa penelitian yang menggunakan analisis klaster non hirarki *K-means* telah dilakukan. Wakhidah (2010) membahas tentang pengklasteran menggunakan algoritma klaster *K-Means*. Kumar dkk (2013) membahas tentang algoritma klaster *K-Means* terbaru dan efisien. Usman dkk (2013) meneliti tentang peningkatan algoritma *K-Means* untuk mendapatkan *centroid* awal. Hasil pengklasteran dengan metode *K-Means* sangat bergantung pada nilai pusat awal klaster (*centroid*) yang diberikan. Pemberian nilai awal yang berbeda bisa menghasilkan klaster yang berbeda (Lathifaturrahmah, 2010). Berdasarkan uraian singkat diatas, penulis tertarik melakukan penelitian mengenai penentuan awal keanggotaan pada analisis klaster non hirarki *K-means*.

## 2. ANALISIS KLASTER NON HIRARKI (*K-MEANS*)

### A. Analisa Klaster

Analisis Klaster merupakan salah satu teknik dalam analisis multivariat yang mempunyai tujuan utama untuk mengelompokkan objek-objek pengamatan menjadi beberapa kelompok berdasarkan karakteristik yang dimilikinya. Analisis klaster mengelompokkan objek-objek sehingga setiap objek yang paling dekat kesamaannya dengan objek lain berada dalam kelompok yang sama, serta mempunyai kemiripan satu dengan yang lain (Johnson & Wichern, 2002). Analisis klaster dapat dibagi menjadi dua langkah dasar yaitu :

1. Pemilihan ukuran kedekatan  
Memeriksa setiap pasangan pengamatan (objek) untuk kemiripan nilainya. Sebuah kemiripan (*proximity*) ukuran didefinisikan untuk mengukur "kedekatan" dari objek. Semakin dekat maka semakin homogen.
2. Pemilihan algoritma kelompok yang membangun  
Berdasarkan kedekatan mengukur obyek ditugaskan untuk kelompok sehingga perbedaan antara kelompok menjadi besar dan pengamatan dalam kelompok menjadi sedekat mungkin (Härdle & Simar, 2007).

Secara logika sebuah klaster dikatakan baik jika mempunyai homogenitas (kesamaan) yang tinggi antar anggota dalam satu klaster (*within cluster*) dan heterogenitas (perbedaan) yang tinggi antar klaster yang satu dengan klaster lainnya (*between cluster*). Analisis klaster harus memenuhi dua asumsi yaitu:

1. Sampel yang diambil harus benar-benar dapat mewakili populasi.
2. Tidak ada multikolinieritas (korelasi antar obyek), jika ada maka nilai multikolinieritasnya kurang dari 0,5 (Nugroho, 2008).

### B. Pengukuran Konsep Kemiripan Antar Jarak Objek

Kemiripan antar objek diukur dengan menggunakan ukuran jarak. Beberapa ukuran jarak yang sering digunakan antara lain jarak *Euclid*, jarak kuadrat *Euclid*, jarak City-block (Manhattan), dan jarak *Chebychev*.

#### 1. Jarak *Euclid*

Jarak *Euclid* biasanya dihitung dari data mentah dan tidak dari data standar. Metode ini memiliki beberapa keuntungan, antara lain jarak dari dua objek tidak dipengaruhi oleh penambahan dari objek baru untuk di

analisis, yang mungkin merupakan pencilan.

Jika  $x' = (x_1, x_2, \dots, x_n)$  dan  $y' = (y_1, y_2, \dots, y_n)$  maka  $d(x, y)$  adalah jarak antara objek  $x$  dengan objek  $y$  (Johnson & Wichern, 2002).

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2}$$
$$d(x, y) = \sqrt{(x - y)'(x - y)} = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

#### 2. Jarak kuadrat *Euclid*

Jarak kuadrat *Euclid* merupakan jumlah kuadrat yang berbeda dari nilai antara dua objek pada seluruh variable.  $d(x, y)$  adalah Jarak antara objek  $x$  dengan objek  $y$  (Fiedling, 2007).

$$\text{jika } x = (x_1, x_2, \dots, x_n) \text{ dan } y = (y_1, y_2, \dots, y_n) \text{ maka}$$
$$d(x, y) = \sum_{i=1}^n (x_i - y_i)^2 \quad (2)$$

untuk  $i = 1, 2, 3, \dots, n$

#### 3. Jarak *city-block* atau *Manhattan*

Jarak ini sering juga disebut jarak *Minkowski*, diberikan oleh jarak antara objek  $x$  dengan objek  $y$  (Johnson & Wichern, 2002).

$$d(x, y) = [\sum_{i=1}^n |x_i - y_i|^m]^{\frac{1}{m}} \quad (3)$$

untuk  $i = 1, 2, 3, \dots, n$

#### 4. Jarak *Chebychev*

Jarak *Chebychev* merupakan jarak yang didasarkan pada nilai perbedaan mutlak maksimum antara dua objek. Pengukuran dengan jarak ini tidak dapat dilakukan untuk data yang bersifat *outlier*.  $d(x, y)$  adalah Jarak antara objek  $x$  dengan objek  $y$  (Mattjik & Sumertajaya, 2011).

$$d(x, y) = \max |x_i - y_i| \quad (4)$$

untuk  $i = 1, 2, 3, \dots, n$

### C. Analisis Klaster Non Hirarki

Analisis klaster dengan metode non hirarki merupakan metode klaster yang menentukan jumlah klaster secara manual (Baroroh, 2012). Teknik analisis klaster non hirarki dirancang untuk mengelompokkan item, bukan variabel, menjadi kumpulan  $K$  klaster. Banyaknya klaster,  $K$ , ditentukan terlebih dahulu untuk memulai prosedur pengklasteran. Metode analisis klaster non hirarki berhubungan dengan *K-means*, ada tiga pendekatan yang digunakan untuk menempatkan masing-masing observasi dalam satu klaster yaitu:

1. ***Sequential Threshold***, Metode *Sequential Threshold* memulai dengan pemilihan satu klaster dan menempatkan semua objek yang berada pada jarak tertentu ke dalamnya sebagai nilai awal klaster. Jika semua objek yang berada pada jarak tertentu telah dimasukkan, klaster yang kedua dipilih dan ditempatkan semua objek yang berjarak tertentu ke dalamnya.

Setelah itu, kluster ketiga dipilih dan proses dilanjutkan seperti yang sebelumnya.

2. **Parallel Threshold**, Metode *Parallel Threshold* berbanding terbalik dengan pendekatan *Sequential Threshold*, metode ini memilih sejumlah kluster secara bersamaan dan menempatkan objek-objek kedalam kluster yang memiliki jarak terdekat kluster awal. Pada saat proses berlangsung, jarak antar kluster awal dapat ditentukan untuk memasukkan beberapa objek ke dalam kluster-kluster. Beberapa variasi pada metode ini, yaitu sisa objek-objek tidak dikelompokkan jika berada di luar jarak tertentu dari sejumlah kluster.
3. **Optimization**, Metode ketiga serupa dengan kedua metode sebelumnya, perbedaannya metode ini memungkinkan untuk menempatkan kembali objek-objek ke dalam kluster yang lebih dekat (Prayudho, 2008).

#### D. K-Means

*K-Means* merupakan metode analisis kluster non hirarki yang dimulai dengan menentukan terlebih dahulu jumlah kluster yang diinginkan. Setelah jumlah kluster diketahui, baru proses kluster dilakukan tanpa mengikuti proses hirarki (Santoso, 2004). *K-Means* banyak digunakan di berbagai bidang karena sederhana dan mudah diimplementasikan. Pengklasteran *K-Means* sangat cocok untuk data dengan ukuran yang besar karena memiliki kecepatan yang lebih tinggi dibandingkan metode hirarki. Hasil pengklasteran mungkin tergantung pada urutan observasi data (Nugroho, 2008).

Algoritma *K-Means* adalah metode klastering berbasis jarak yang membagi data ke dalam sejumlah kluster dan hanya bekerja pada atribut numerik. Secara umum metode analisis kluster *K-means* menggunakan algoritma sebagai berikut:

1. Tentukan  $k$  sebagai jumlah kluster yang di bentuk.  
Untuk menentukan banyaknya kluster  $K$  dilakukan dengan beberapa pertimbangan seperti pertimbangan teoritis dan konseptual yang mungkin diusulkan untuk menentukan berapa banyak kluster.
2. Bangkitkan  $K$  titik pusat kluster (*centroid*) awal secara *random*.  
Penentuan *centroid* awal dilakukan secara *random*/acak dari objek-objek yang tersedia sebanyak  $K$  kluster, kemudian untuk menghitung *centroid* kluster ke- $i$  berikutnya, digunakan rumus sebagai berikut :

$$v = \frac{\sum_{i=1}^n x_i}{n} ; i = 1,2,3,\dots,n \quad (5)$$

Dimana:

$v$  = *centroid* pada kluster  
 $x_i$  = objek ke- $i$   
 $n$  = banyaknya objek

3. Hitung jarak setiap objek ke masing-masing *centroid* dari masing-masing kluster. Untuk menghitung jarak antara objek dengan *centroid* digunakan jarak *euclidian* yang dirumuskan sebagai berikut:

$$d(i, l) = (\sum_{j=1}^p [x(i, j) - \bar{x}(j, l)]^2)^{1/2} \quad (6)$$

Dimana :

$d(i, l)$  = jarak *euclid*  
 $x(i, j)$  = nilai objek ke- $i$  pada variabel ke- $j$   
 $\bar{x}(j, l)$  = rata-rata objek ke- $i$  pada variabel ke- $j$   
 $i$  = banyaknya objek  
 $j$  = banyaknya variabel  
 $l$  = banyaknya kluster

4. Alokasikan masing-masing objek ke dalam *centroid* yang paling terdekat. Untuk melakukan pengalokasian objek kedalam masing-masing kluster pada saat iterasi secara umum dapat dilakukan dengan dua cara yaitu dengan *hard k-means*, dimana secara tegas setiap objek dinyatakan sebagai anggota kluster dengan mengukur jarak kedekatan terhadap titik pusat kluster tersebut, cara lain dapat dilakukan dengan *fuzzy C-Means*.
5. Lakukan iterasi, kemudian tentukan posisi *centroid* baru dengan menggunakan persamaan (2.7).
6. Ulangi langkah 3 jika posisi *centroid* baru tidak sama.  
Pengecekan konvergensi dilakukan dengan membandingkan matriks *group assignment* (matrik kelompok yang disimbolkan dengan  $G^m$  dimana  $m$  merupakan banyaknya iterasi) pada iterasi sebelumnya dengan matrik *group assignment* pada iterasi yang sedang berjalan. Jika hasilnya sama maka algoritma *K-means* sudah konvergen, tetapi jika berbeda maka belum konvergen sehingga perlu dilakukan iterasi berikutnya.

Berdasarkan cara kerjanya, algoritma *K-Means* memiliki kelebihan dan kekurangan antara lain sebagai berikut:

##### A. Kelebihan *K-Means*

1. Algoritma *K-Means* adalah algoritma klasik untuk menyelesaikan masalah pengklasteran dikarenakan algoritma ini relatif sederhana dan cepat.
2. Algoritma *K-Means* memberikan hasil yang relatif baik pada kluster konveks (Zhang & Fang, 2013).
3. Algoritma *K-Means* memiliki ketelitian yang cukup tinggi terhadap ukuran objek, sehingga algoritma ini relatif lebih terukur dan efisien untuk mengolah data dalam jumlah besar.
4. Algoritma *K-Means* tidak berpengaruh pada urutan objek (Simamora, 2005).

#### B. Kekurangan *K-Means*

1. Sangat sensitif pada pembangkitan titik pusat awal secara *random* (acak).
2. Memungkinkan suatu kluster tidak memiliki anggota
3. Hasil pengklasteran bersifat tidak unik (selalu berubah-ubah) terkadang bagus terkadang tidak.
4. Sulit untuk mencapai *global optimum*.
5. Menentukan banyaknya jumlah kluster sebelum mengetahui jumlah kluster yang optimal (Lathifaturrahmah, 2010).
6. Tidak jelas berapa banyak kluster *K* yang terbaik.
7. Hanya bekerja pada atribut numerik.
8. Hanya dapat digunakan dalam data yang *mean*-nya dapat ditentukan
9. Tidak mampu menangani data yang mempunyai penyimpangan-penyimpangan (*noisy data* dan *outlier*) (Andayani, 2007).

#### E. Indeks Validitas

Indeks validitas merupakan suatu ukuran yang digunakan untuk menentukan jumlah kelompok yang optimal. Untuk menentukan jumlah kluster yang paling optimal dan dapat memvalidasi apakah partisi yang diterapkan dalam proses klasterisasi sesuai dengan data, maka digunakan indeks pengukuran kluster validasi. Validitas dalam *K-Means* ditentukan dengan banyaknya kelompok optimum melalui perhitungan indeks validitas. Salah satu indeks validitas yang digunakan dalam *K-Means* adalah indeks *XB*.

Formula indeks *XB* sebagai berikut:

$$XB(c) = \frac{\sum_{i=1}^n \sum_{k=1}^c [\sum_{j=1}^m (x_{ij} - v_{kj})^2] (U_{ik})^w}{N \min \|v_{kj}^1, v_{kj}^t\|^2} \quad (7)$$

Dimana :

- $c$  = jumlah kluster
- $N$  = jumlah objek yang akan dikelompokkan
- $U_{ik}$  = derajat keanggotaan data ke-*i* pada kluster ke-*k*
- $\|v_{kj}^1, v_{kj}^t\|^2$  = jarak antara pusat kluster  $v_{kj}^1$  dan  $v_{kj}^t$
- $t$  = banyaknya iterasi atau perulangan
- $\min \|v_{kj}^1, v_{kj}^t\|^2$  = jarak minimum antara pusat kluster  $v_{kj}^1$  dan  $v_{kj}^t$

### 3. PENENTUAN AWAL KEANGGOTAAN ANALISIS KLUSTER NON HIRARKI (*K-MEANS*)

#### A. Komponen Utama

Dalam metode non hirarki (*K-Means*) data dibagi menjadi *K* partisi dimana setiap partisi mewakili sebuah kluster. Dasar pengklasteran dalam metode ini adalah menempatkan objek berdasarkan rata-rata (mean) kluster terdekat. Oleh karena itu, metode ini bertujuan untuk meminimumkan *error* akibat partisi *n* objek ke dalam *k* kluster (Nuningsih, 2010). Algoritma *K-Means* dalam proses pengklasterannya memerlukan 3 komponen yaitu:

##### 1. Jumlah Kluster

Untuk menentukan banyaknya kluster *K* dilakukan dengan beberapa pertimbangan seperti pertimbangan teoritis dan konseptual yang mungkin diusulkan untuk menentukan berapa banyak kluster. Jumlah kluster *K* juga dapat ditentukan melalui pendekatan hirarki. Namun, perlu diperhatikan bahwa tidak terdapat aturan khusus dalam menentukan jumlah kluster *K*, terkadang jumlah kluster yang diinginkan tergantung pada subjektif seseorang.

##### 2. Kluster Awal

Kluster awal yang dipilih berkaitan dengan penentuan pusat kluster awal (*centroid* awal). Dalam hal ini, terdapat beberapa pendapat dalam memilih kluster awal untuk metode *K-Means* sebagai berikut:

- a. Berdasarkan Hartigan (1975), pemilihan kluster awal dapat

ditentukan berdasarkan interval dari jumlah setiap observasi.

- b. Berdasarkan Rencher (2002), pemilihan kluster awal dapat ditentukan melalui pendekatan salah satu metode hirarki
- c. Berdasarkan Teknomo (2007), pemilihan kluster awal dapat secara acak dari semua observasi.

Oleh karena adanya pemilihan kluster awal yang berbeda ini maka kemungkinan besar solusi kluster yang dihasilkan akan berbeda pula.

### 3. Ukuran Jarak

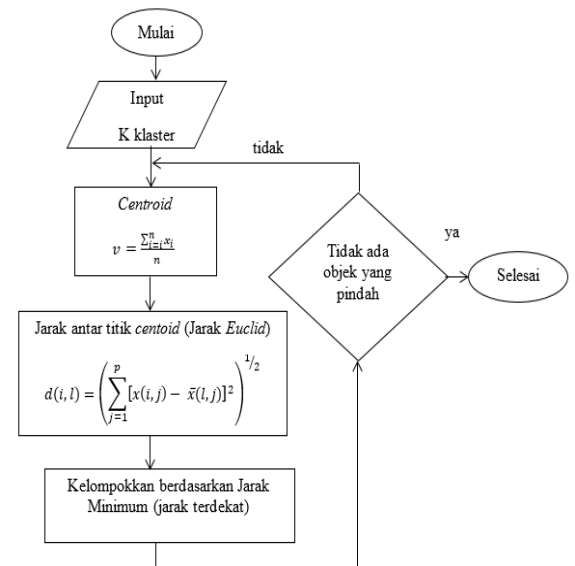
Dalam hal ini, ukuran jarak digunakan untuk menempatkan observasi ke dalam kluster berdasarkan centroid terdekat. Ukuran jarak yang digunakan dalam metode *K-Means* adalah jarak *euclid* (Nuningsih, 2010).

## B. Penentuan Awal Keanggotaan Analisis Kluster Non Hirarki (*K-Means*)

Kinerja algoritma *K-Means* dapat dipengaruhi oleh nilai yang dipilih dari *K*. Oleh karena itu, ditetapkan sebuah nilai *K*. Hal ini penting untuk nilai *K* yang dianggap cukup besar, untuk mencerminkan karakteristik khusus dari himpunan data. Nilai *K* yang dipilih harus secara signifikan lebih kecil dari jumlah objek dalam himpunan data, yang merupakan tujuan utama untuk melakukan pengelompokan data (Pham dkk, 2004). Pada penelitian ini, kita akan menentukan penentuan awal keanggotaan analisis kluster non hirarki *K-Means* yaitu *K* kluster yang dapat dibentuk dari *n* data dan membagi objek menjadi *K* himpunan dengan memperhatikan semua kemungkinan. Berikut langkah-langkah penentuan awal keanggotaan pada analisis kluster non hirarki *K-Means*:

1. Tentukan nilai *K* sebagai jumlah kluster yang dibentuk. Pada penelitian ini, semua kemungkinan kluster akan digunakan dimana nilai harus *K* lebih kecil dari jumlah objek dalam himpunan data.
2. Bangkitkan *K* titik pusat kluster (*centroid*) awal secara acak, dimana kita bisa mencoba semua kemungkinan centroid dari *n* data.
3. Hitung jarak setiap objek ke masing-masing *centroid* dari masing-masing kluster dengan menggunakan jarak *euclidian* pada persamaan (2.1). Kemudian buat matriks jarak yang disimbolkan dengan  $D^m$ , dimana *m* adalah iterasi proses pengklasteran.

4. Pengelompokan objek, untuk menentukan anggota kluster kita dapat memperhitungkan jarak minimum objek. Nilai yang diperoleh dalam keanggotaan data pada matriks jarak adalah 0 atau 1, dimana nilai 1 untuk data yang dialokasikan ke kluster dan nilai 0 untuk data yang dialokasikan ke kluster lain. Sama halnya seperti langkah 3, untuk mempermudah pengelompokan maka dibuat sebuah matriks yaitu matriks kelompok atau *group* yang dilambangkan dengan  $G^m$ , dimana *m* adalah iterasi proses pengklasteran.
5. Kembali ke langkah 2, lakukan pengulangan hingga nilai *centroid* yang dihasilkan tetap dan anggota kluster tidak berpindah ke kluster lain. Secara flowchart algoritma *K-Means* digambar sebagai berikut (Wakhidah, 2010):



Gambar 1. Flowchart Algoritma *K-Means*

## 4. TELADAN PENERAPAN DAN PEMBAHASAN

### A. Teladan Penerapan

Untuk lebih memahami materi mengenai penentuan awal keanggotaan pada analisis kluster non hirarki *K-Means* diberikan teladan penerapan yang diperoleh dari buku berjudul “Applied Multivariate Statistical Second Edition” karangan Johnson dan Winchern (2002: 741) berupa data 2 variabel  $X_1$  dan  $X_2$  dengan empat item *A, B, C, D*. Adapun data tersebut yaitu:

Tabel 1 Data Teladan Penerapan

Item	Observasi	
	$x_1$	$x_2$
A	5	4
B	1	-2
C	-1	1
D	3	1

### B. Pembahasan

Dalam menyelesaikan persoalan pada teladan penerapan, hal pertama yang kita lakukan adalah menentukan nilai  $K$  sebagai jumlah kluster yang dibentuk. Dengan memperhatikan semua kemungkinan pada data Tabel.1 kita dapat mengelompokkan item menjadi 2 dan 3 kluster.

Setelah penentuan nilai  $K$ , kita tentukan nilai kluster awal secara acak dengan memperhatikan semua kemungkinan dari empat item pada Tabel.1 dimana untuk  $K=2$  yaitu (AB dan CD), (AC dan BD), (AD dan BC), (A dan BCD), (B dan ACD), (C dan ABD) dan (D dan ABC). Untuk  $K=3$  yaitu (A, B dan CD), (A, BC dan D), (AB, C dan D), (AC, B dan D), (AD, B dan C) dan (A, BD dan C).

Selanjutnya, setelah penentuan kluster awal maka dapat dihitung nilai *centroid*, jarak antar titik terhadap *centroid*, mengelompokkan data sesuai jarak minimum jika tidak terjadi perpindahan kluster proses berhenti dan jika terjadi perpindahan maka dihitung kembali nilai *centroid* baru dari kemungkinan kluster awal yang telah diperoleh.

Dengan melakukan proses pengklasteran seperti diatas diperoleh hasil akhir sebagai berikut:

Untuk  $K=2$

Tabel 1. Hasil Akhir Pengklasteran Untuk  $K=2$

Kluster Awal	Kluster Akhir
AB dan CD	AD dan BC
AC dan BD	ACD dan B
	A dan BCD
	AD dan BC
AD dan BC	AD dan BC
A dan BCD	A dan BCD
B dan ACD	B dan ACD
C dan ABD	BC dan AD
D dan ABC	A dan BCD

Untuk  $K=3$

Tabel 2. Hasil Akhir Pengklasteran Untuk  $K=2$

Kluster Awal	Kluster Akhir
A, B dan CD	A, B dan CD
A, BC dan D	A, BC dan D
AB, C dan D	A, BC dan D
	AD, C dan B
	A, C dan BD
AC, B dan D	AC, B dan D
AD, B dan C	AD, B dan C
A, BD dan C	A, BD dan C

## 5. KESIMPULAN DAN SARAN

### A. Kesimpulan

Berdasarkan penelitian yang dilakukan tentang penentuan awal keanggotaan analisis kluster non hirarki  $K$ -Means dapat diambil kesimpulan sebagai berikut:

1. Dalam menentukan awal keanggotaan pada analisis kluster non- hirarki  $K$ -Means yaitu jumlah  $K$ , secara signifikan nilai  $K$  yang dipilih harus lebih kecil dari jumlah objek dalam himpunan data. Jumlah kluster  $K$  juga dapat ditentukan melalui pendekatan hirarki. Namun, perlu diperhatikan bahwa tidak terdapat aturan khusus dalam menentukan jumlah kluster  $K$ , terkadang jumlah kluster yang diinginkan tergantung pada subjektif seseorang.
2. Dari teladan penerapan pada penentuan awal keanggotaan analisis kluster non hirarki  $K$ -Means dapat disimpulkan bahwa algoritma  $K$ -Means tidak selalu menghasilkan hasil pengklasteran yang sama. Hal ini dikarenakan adanya unsur pengacakan pada titik pusat kluster dan jumlah sampel yang sedikit yaitu  $n < 30$ .

### B. Saran

Dalam penulisan skripsi ini, penulis hanya membahas masalah penentuan awal keanggotaan analisis kluster non hirarki  $K$ -Means pada data 2 dimensi dengan mencoba semua kemungkinan  $K$ . Untuk penelitian lebih lanjut, sebaiknya penentuan awal keanggotaan analisis kluster non hirarki  $K$ -Means diaplikasikan pada data lebih dari 2 dimensi dengan mencoba semua kemungkinan  $K$ , namun dalam pengukuran kemiripan antar objek menggunakan 2 jarak yaitu jarak *city-block* (*manhattan*) dan jarak *euclid* sebagai perbandingan.

## 6. DAFTAR PUSTAKA

- Agusta, Y. 2007. *K-Means – Penerapan, Permasalahan dan Metode Terkait*. Jurnal Sistem dan Informatika Vol.3 (Februari 2007), 47-60.
- Andayani, S. 2007. Pembentukan *cluster* dalam *Knowledge Discovery in Database* dengan Algoritma *K-Means*. SEMNAS Matematika dan Pendidikan Matematika. Jurdik Matematika FMIPA UNY Yogyakarta pada tanggal 24 Nopember 2007. <http://staff.uny.ac.id/sites.pdf>. (Diakses tanggal 20 Mei 2013).
- Baroroh, A. 2012. *Analisis Multivariat dan Time Series dengan SPSS 21*. Tangerang Selatan: PT. ELEK MEDIA KOMPUTINDO.
- Ediyanto, Sari, M.N., dan Satyahadewi, N. 2013. Pengklasifikasian Karakteristik Dengan Metode *K-Means Cluster Analysis*. Buletin Ilmiah Mat. Stat. dan Terapannya(Bimaster) Volume 02, No.2 (2013), hal 133-136. <http://download.portalgaruda.org/article.php?article=79314&val=2316>. (Diakses tanggal 19 April 2013).
- Härdle, W and Simar, L. 2007. *Applied Multivariate Statistical Analysis Second Edition*. Springer: Berlin and Louvain la Neuve.
- Heryandi, A. 2009. Kajian Analisis Clustering Dan Penerapannya Pada Sistem Pendukung Keputusan. *Tesis*. Program Studi Magister Informatika. ITB.
- Johnson, R. A and D, W. Wichern. 2002. *Applied Multivariate Statistical Analysis, Fifth Edition*. Pearson Education. United States Of America.
- Kumar, G. K., Chary, T. B., and Premchand, P. 2013. A New and Efficient K-Means Clustering Algorithm. *Volume 3, Issue 11, November 2013*. <http://www.cs.umd.edu/~mount/Projects/KMeans/pami02.pdf>. (Diakses tanggal 19 April 2013).
- Lathifaturrahmah. 2010. Perbandingan Hasil Penggerombolan Metode *K-Means*, *Fuzzy K-Means*, dan *Two Step Cluster*. Tesis, Departemen Matematika Sekolah Pasca Sarjana Institut Pertanian Bogor: Bogor. Tidak Dipublikasikan. <http://repository.ipb.ac.id/handle/123456789/26950>. (Diakses tanggal 22 April 2013)
- Martiana, E., Mubtada'i N.R., dan Purnomo.E. 2012. Penggunaan Metode Pengklasteran Untuk Menentukan Bidang Tugas Akhir Mahasiswa Teknik Informatika. <http://repo.eepis-its.edu/130/>. (Diakses tanggal 22 April 2013)
- Mattjik, A. A dan I. M. Sumertajaya. 2011. Sidik Peubah Ganda Dengan menggunakan SAS. IPB: Bandung.
- Nugroho, S. 2008. *Statistika Multivariat Terapan*. Unib Press: Bengkulu.
- Nuningsih, Sri. 2010. *K-Means Clustering: Studi Kasus Pada Data Pengujian Kualitas Susu Di Koperasi Peternakan Bandung Selatan*. *Skripsi*. Fakultas Matematika Dan Ilmu Pengetahuan Alam. UPI. <http://a-research.upi.edu/skripsiview.php?start=8112>. (Diakses tanggal 1 Oktober 2013)
- Pham. D. T., Dimov. S. S., and Nguyen. C.D. 2004. Selection of *K* in *K-means Clustering*. *Manufacturing Engineering Centre, Cardiff University, Cardiff, UK*. <http://www.ee.columbia.edu/~dpwe/papers/PhamDN05-kmeans.pdf>. (Diakses tanggal 1 Mei 2013).
- Prayudho, B.J., 2008. Analisis Cluster. <http://prayudho.wordpress.com>. (Diakses tanggal 19 April 2013).
- Rencher, A.C., 2002. *Methods Of Multivariate Analysis Second Edition*. Wiley Series In Probability and Mathematical Statistics: Canada.
- Santoso, S. 2010. *Statistika Multivariat Konsep & Aplikasi dengan SPSS*. Jakarta: PT. Elex Media Komputindo.
- Simamora B. 2005. *Analisis Multivariat Pemasaran*. PT. Gramedia Pustaka Utama: Jakarta.
- Supranto, J. 2004. *Analisis Multivariat Arti dan Interpretasi*. Rineka Cipta: Jakarta.
- Usman, G., Ahmad, U., dan Ahmad, M. 2013. Improved K-Means Clustering Algorithm by Getting Initial Cenroids. *World Applied Sciences Journal 27 (4): 543-551, 2013*. [http://www.idosi.org/wasj/wasj27\(4\)13/21.pdf](http://www.idosi.org/wasj/wasj27(4)13/21.pdf). (Diakses tanggal 1 Mei 2013)
- Wakhidah, N. 2010. Clustering Menggunakan *K-Means Algorithm (K-Means Algorithm Clustering)*. *Tr@nsForMat!ka Vol 8 No 1*.
- Zhang, C and Fang, Z. 2013. An Improved *K-Means Clustering Algorithm*. *Journal of Information & Computational Science 10: 1 (2013) 193–199*.

## **PENERAPAN EMPIRICAL MODE DECOMPOSITION PADA PERGERAKAN HARGA CABAI DI PROVINSI BENGKULU**

**Sarah Tiur Mauli<sup>1</sup>, Prof, Ir. Sigit Nugroho, PhD<sup>2</sup>, Jose Rizal, S. Si., M. Si.<sup>2</sup>,**

<sup>1</sup> Alumni Jurusan Matematika, Fakultas MIPA, Universitas Bengkulu

<sup>2</sup>Dosen Jurusan Matematika, Fakultas MIPA, Universitas Bengkulu

**Cg.sarah@yahoo.co.id**

Metode *Empirical Mode Decomposition* (EMD) merupakan pendekatan alternatif analisis harga cabai melalui proses dekomposisi data menjadi beberapa *intrinsic modefunction* (IMF) dan residu. Metode EMD mampu bekerja pada kondisi data yang bersifat nonlinear dan nonstasioner sehingga sesuai dengan karakteristik harga cabai yang tidak stabil antar musim dan tahun. Penelitian ini dilakukan terhadap perkembangan harga cabai harian dan mingguan di Provinsi Bengkulu. Berdasarkan hasil EMD, tiga IMF yang memiliki kontribusi terbesar terhadap volatilitas harga cabai harian ialah IMF 1, IMF 2 dan IMF 4 serta IMF 1, IMF 3, dan IMF 4 pada data harga cabai mingguan. Komponen tren memberikan kontribusi dominan dengan nilai 89,94% untuk data harga harian dan 83,96% untuk data harga mingguan. Artinya data harga harian memperlihatkan komponen kecenderungan berfrekuensi tinggi atau dengan kata lain data harga cabai harian memberikan kontribusi kecenderungan terbaik sehingga memberikan pengaruh lebih besar.

**Kata Kunci:** *EMD, Time Series, IMF, Volatilitas Harga..*

### **PENDAHULUAN**

Hortikultura adalah tanaman budidaya yang menunjang pemenuhan gizi masyarakat sebagai sumber vitamin, mineral, protein, dan karbohidrat. Permintaan pasar domestik maupun pasar internasional terhadap komoditas hortikultura di masa mendatang diperkirakan akan meningkat seiring dengan peningkatan jumlah penduduk dan tingkat pendapatan (Anonim, 2007).

Komoditas hortikultura merupakan komoditas yang sangat penting karena jenis komoditas ini merupakan kebutuhan pokok manusia, yang setiap saat selalu tersedia dalam jumlah cukup dengan mutu yang layak, aman dikonsumsi dan dengan harga

yangterjangkau oleh masyarakat (Anonim, 2010).

Cabai merupakan salah satu komoditas pertanian yang masuk dalam subsector hortikultura yang berperan penting dalam mendukung perekonomian nasional, khususnya dalam meningkatkan pendapatan masyarakat. Kontribusi subsector hortikultura terhadap Produk Domestik Bruto nasional/PDB (berdasarkan harga yang berlaku) pada tahun 2005 mencapai Rp 61.792,44 triliun, dan meningkat menjadi Rp 68.640,39 triliun pada 2006, Rp 74.768 triliun pada 2007 dan target menjadi Rp 78.292 triliun pada 2008.

Metode *Empirical Mode Decomposition* (EMD) diperkenalkan oleh Huang *et al.*(1998). Metode ini banyakdigunakan pada bidang ilmu geofisika dan biomedis. EMD ialah teknik analisis yangempiris, langsung dan adaptif terhadap pemrosesan data, khususnya data yang bersifat tidak linear dan tidak stasioner.Hal ini sesuai dengan karakteristik harga cabai yang memiliki ketidakstabilan antar musim dan tahun. Prinsip EMD ialah menguraikan data runtun waktu menjadi sejumlah *Intrinsic Mode Function* (IMF) yang bebas dan cenderung periodik berdasarkan skala karakteristik lokal (Zhang *et al.*, 2008). Kelemahan dari metode EMD ini adalah dihasilkannya *mode mixing* (modus campuran) sehingga Wu dan Huang (2004) mengembangkan EMD melalui konsep *ensemble* agar skala karakteristik yang dihasilkan dalam IMF menjadi lebih natural dengan menambahkan serangkaian *white noise* pada data.

Penelitian kali ini akan mengaplikasikan metode EMD untuk menganalisis perilaku harga cabai mingguan dan harian di Provinsi Bengkulu selama dua tahun terakhir. Analisis karakteristik harga cabai mingguan dan harian tersebut akan difokuskan pada tahap penguraian serangkaian IMF yang dihasilkan algoritma EMD.

Metode EMD juga mampu memberikan hasil analisis dari kontribusi kecenderungan terbaik dari masing-masing selang waktu pengamatan. Hal ini dapat dijadikan sebagai bahan pertimbangan selang waktu terbaik dalam melakukan peramalan data berikutnya.

### **Rumusan masalah**

Dapat dirumuskan masalah dalam penelitian ini sebagai berikut:

1. Bagaimana aplikasi dari metode EMD dalam menganalisis perilaku harga cabai berdasarkan data cabai mingguan dan harian dari Januari 2014 sampai Pebruari 2016 ?
2. Interval waktu manakah (harian atau mingguan) yang memberikan kontribusi kecenderungan terbaik yang dapat digunakan untuk melakukan peramalan data berikutnya ?

### **Tujuan Penulisan**

Berdasarkan rumusan masalah diatas, maka tujuan penelitian ini ialah untuk :

1. Mendapatkan sebuah cara dalam mengaplikasikan pola deret waktu dari perkembangan harga cabai periode mingguan dan harian dari Januari 2014 sampai Pebruari 2016 dengan pendekatan EMD.
2. Untuk melihat Interval waktu yang memberikan kontribusi kecenderungan terbaik yang dapat digunakan untuk melakukan peramalan data berikutnya.

## **METODE PENELITIAN**

### **Objek Penelitian**

Objek penelitian dalam skripsi ini adalah harga cabai merah di Provinsi Bengkulu. Dimana akan dilakukan analisis pola deret waktu dari perkembangan harga cabai mingguan dan harian. Analisis yang digunakan adalah deret waktu dengan metode *Empirical Mode Decomposition*.

### **Jenis Penelitian**

Penelitian ini dikategorikan penelitian lapangan (*field research*) atau studi kasus yaitu penelitian yang mengumpulkan dokumen objek di Dinas Perdagangan Provinsi Bengkulu.

### **Jenis dan Sumber Data**

Data yang digunakan dalam penelitian ini merupakan data sekunder. Data sekunder adalah data yang didapat dari dokumen yang dimiliki oleh Dinas Perdagangan Provinsi Bengkulu.

Data yang diperlukan dalam penelitian ini adalah data harga cabai dalam periode mingguan dan harian dari Januari 2014 – Pebruari 2016. Pada penelitian ini diambil data selama 2 tahun. Jenis datanya adalah runtun waktu. Adapun cara pengambilan data yang dilakukan dengan menggunakan data yang tersedia.

### **Analisis Data**

Metode yang digunakan adalah *Empirical Mode Decomposition* (EMD). Metode EMD merupakan metode adaptif yang dirancang untuk mempresentasikan data tidak stasioner dan tidak linier sehingga bertujuan untuk memisahkan data menjadi beberapa bagian data (IMF) melalui proses dekomposisi.

Setelah data yang akan digunakan terkumpul, maka tahapan selanjutnya adalah pemeriksaan data deret waktu yang akan digunakan agar tidak terdapat komponen data yang kosong melalui penanganan *missing value*. Selanjutnya tahapan yang digunakan dalam penelitian ini adalah :

1. Membuat plot data input harga cabai mingguan dan harian.
2. Mengekstrak kumpulan IMF dan residu berdasarkan algoitma EMD.

3. Mengetahui kontribusi masing-masing IMF yang dihasilkan terhadap data.
4. Mengetahui kontribusi komponen berfrekuensi tinggi dan komponen berfrekuensi rendah yang dihasilkan terhadap data input harga cabai mingguan dan harian.

Dalam penelitian ini digunakan uji stasioneritas *Phillip-Perron* (PP) untuk melihat kestasioneran data dengan menggunakan software *eviews 8.0*. Kemudian untuk mencari IMF dan nilai residu akan dibantu software *R.3.1.1* dan *Ms.Excel*.

## HASIL DAN PEMBAHASAN

Data yang digunakan dalam penelitian ini adalah data sekunder yang diambil dari Dinas Perdagangan Provinsi Bengkulu. Data ini adalah data harga cabai di Provinsi Bengkulu. Pengumpulan data dilakukan dalam periode dari bulan Januari 2014 sampai dengan Pebruari 2016 atau sebanyak 109 data untuk data mingguan dan 765 data untuk data harian. Data pertama yang digunakan adalah data harga cabai harian periode Januari 2014 sampai dengan Pebruari 2016 yang ada sebanyak 765 data harga cabai harian. Kemudian data kedua yang digunakan adalah data harga cabai mingguan pada periode yang sama yaitu Januari 2014 sampai dengan Pebruari 2016 yang ada sebanyak 109 data.

### Kestasioneran Data

Stasioner artinya tidak terdapat pertumbuhan atau penurunan pada data. Akar unit merupakan salah satu cara untuk menguji kestasioneran data, dalam artian data dikatakan stasioner apabila tidak mengandung akar unit. Untuk kestasioneran data yang digunakan dalam penelitian ini adalah dengan menggunakan uji akar unit *Phillips-Perron*. Hipotesis pengujian dari akar unit *Phillips-Perron* adalah sebagai berikut :

: (data mengandung akar unit atau data non stasioner)

: (data tidak mengandung akar unit atau data stasioner)

Tolak jika atau jika

**Tabel. 4.3.1. Uji *Phillips-Perron* Harga Cabai Harian**

	Adj. t-Stat	Prob.*
Phillips-Perron test statistic	-2.274823	0.1805

Dari Tabel 4.3.1 dapat disimpulkan bahwa diterima karena nilai probabilitas sebesar

0,185 yang masih lebih besar dari pada nilai kritis = 0,05. Sehingga dapat disimpulkan bahwa pada taraf 5% diterima, artinya data mempunyai akar unit dan data dapat dikatakan data bersifat tidak stasioner. Sedangkan untuk harga cabai mingguan dapat dilihat pada Tabel 4.3.2 dibawah ini.

**Tabel. 4.3.2. Uji *Phillips-Perron* Harga Cabai Mingguan**

	Adj. t-Stat	Prob.*
Phillips-Perron test statistic	-2.306484	0.1718

Dari Tabel 4.3.2 juga dapat disimpulkan bahwa  $H_0$  diterima karena nilai probabilitas sebesar 0,1718 yang masih lebih besar dari pada nilai kritis = 0,05. Sehingga dapat disimpulkan bahwa pada taraf 5% diterima, artinya data mempunyai akar unit dan data dapat dikatakan data bersifat tidak stasioner. Sehingga harga cabai harian dan mingguan sudah memenuhi syarat untuk dilakukan proses dekomposisi (EMD), namun sebelum melakukan dekomposisi terlebih dahulu dilakukan pengujian kelinieran data.

**Kelinieran Data**

Pengujian linieritas berfungsi untuk mengetahui apakah data bersifat linier atau tidak linier. Dimana sifat ketidaklinieran data merupakan salah satu syarat yang akan digunakan untuk melakukan proses pendekomposisian data. Hasil pada pengujian linieritas yang menunjukkan ketidaklinieran dapat dilihat pada tabel dibawah ini.

**Tabel. 4.4.1. Uji *Ramsey RESET* Harga Cabai Harian**

	Value	df	Probability
t-statistic	56.79104	763	0.0000
F-statistic	3225.222	-1,763	0.0000
Likelihood ratio	1265.19	1	0.0000

Berdasarkan Tabel 4.4.1 dapat disimpulkan bahwa diterima karena nilai kritis =

0,05 lebih besar dari nilai *p.value* yaitu 0,0000. Sehingga dapat disimpulkan bahwa pada taraf 5% diterima, artinya data tidak linier. Selanjutnya untuk data cabai mingguan dapat dilihat pada tabel berikut :

**Tabel. 4.4.2. Uji Ramsey RESET Harga Cabai Mingguan**

	Value	df	Probability
t-statistic	21.59729	107	0.0000
F-statistic	466.4430	(1, 107)	0.0000
Likelihood ratio	182.9924	1	0.0000

Berdasarkan Tabel 4.4.2 dapat dilihat bahwa pada nilai kritis = 0,05 lebih besar dari nilai *p.value* yaitu 0,000. Sehingga dapat disimpulkan bahwa diterima yang artinya data tidak linier. Setelah diuji kestasioneran dan kelinieran data, diketahui bahwa data layak untuk dilanjutkan ketahap selanjutnya yakni estimasi EMD.

### Estimasi EMD

**Tabel 4.5.1. Rataan, Korelasi, Ragam dan Tren Harga Cabai Harian**

Hasil EMD	Jumlah Puncak	Rataan	Korelasi Pearson	Ragam	Presentasi Rasio Ragam
Data awal				362072573	
IMF 1	158	31,04	0,21	8143080,84	2,24 %
IMF 2	63	50,36	0,36	2695595,17	0,74 %
IMF 3	24	53,27	0,09	525429,17	0,145%
IMF 4	8	32,74	0,12	367852,40	0,10 %
IMF 5	3	24,2	0,03	61637,53	0,017%
Residu				325648072,2	89,94 %
Jumlah					93,182%

Pada Tabel 4.5.1 dapat dilihat dari rata-rata periode yang terdapat pada tiap-tiap IMF. Hubungan dari tiap-tiap IMF dapat ditinjau berdasarkan korelasi terhadap data input, ragam, dan persentase rasio ragam terhadap data input.

Berdasarkan Tabel 4.5.1 IMF yang memiliki korelasi di atas 0.1 ialah IMF 1, IMF 2 dan IMF 4. Pola ini mencerminkan bahwa IMF dengan indeks relatif tinggi memiliki pergerakan harga cabai (naik ataupun turun) yang bertahan untuk waktu yang lama

sebelum arah pergerakannya berubah, sehingga cukup erat hubungannya dengan kondisi data awal dengan perubahan besar lebih dipengaruhi oleh efek kejadian-kejadian insidental di tahun-tahun tertentu. Sementara itu, IMF dengan indeks kecil yang memiliki rata-rata periode pendek cenderung sangat fluktuatif dan berlawanan dengan arah gerak harga secara umum. Oleh karena itu, golongan IMF ini tidak memiliki hubungan yang erat. Selanjutnya residual memiliki korelasi tertinggi dengan data. Hubungan yang erat antara residu dan data awal disebabkan oleh arah pergerakan residu yang monoton naik. Hal ini sesuai dengan arah data awal di sebagian besar observasinya. Sementara itu, 89,94% sumber volatilitas harga cabai harian ini berasal dari residual. Hal ini menjelaskan bahwa residual dapat dilihat sebagai tren pergerakan harga residual cabai untuk jangka waktu yang panjang.

**Tabel 4.5.2. Rataan, Korelasi, Ragam dan Tren Harga Cabai Mingguan**

Hasil EMD	Jumlah Puncak	Rataan	Korelasi Pearson	Ragam	Presentasi Rasio Ragam
Data awal				357421761,4	
IMF 1	33	49,35	0,70%	22546120,67	6,3 %
IMF 2	11	56,65	0,30%	6245388,64	1,74%
IMF 3	4	18,86	0,08%	6270436,38	1,75%
IMF 4	2	194,14	0,15%	7476753,91	2,09%
Residu				300091310,5	83,96%
Jumlah					95,84%

Berdasarkan Tabel 4.5.2 dapat diketahui bahwa IMF 1 merupakan IMF dengan kontribusi terbesar terhadap perubahan harga mingguan yakni sebanyak 6,3%. Selanjutnya, IMF 3 dan IMF 4 secara berurutan berkontribusi sebesar 1,75% dan 2,09%. Kemudian residu memberikan kontribusi terbesar yaitu 83,96%. Secara umum, hasil dekomposisi pada data mingguan dan harian mengindikasikan kontribusi yang besar dari IMF. Kemudian Tabel 4.5.1 dan Tabel 4.5.2, total ragam yang mampu dijelaskan oleh semua IMF dan residu hanya 93,182% dan 95,84% saja. Keduanya belum mampu menjelaskan seluruh ragam data input. Fenomena ini dijelaskan oleh Peel *et al.* (2005) sebagai akibat dari adanya kombinasi dari kesalahan pembulatan, nonlinearitas dari data input dan ragam yang ditimbulkan proses interpolasi *cubic spline*.

## **PENUTUP**

## Kesimpulan

Berdasarkan hasil peneitian dapat disimpulkan bahwa :

1. Data harga cabai harian dan mingguan di Provinsi Bengkulu dapat diaplikasikan pada metode *Empirical Mode Decomposition* (EMD) karena data tersebut memiliki karakteristik data yang non linier dan non stasioner.
2. Berdasarkan hasil EMD tiga IMF yang memiliki kontribusi terbesar terhadap volatilitas harga cabai harian ialah IMF 1, IMF 2 dan IMF 4. Sedangkan yang memiliki kontribusi terbesar terhadap volatilitas harga cabai mingguan ialah IMF 1, IMF 3 dan IMF 4. Komponen kecenderungan memberikan kontribusi dominan dengan nilai masing-masing 89,94% untuk data harga harian dan 83,96% untuk data harga mingguan. Artinya data harga harian memperlihatkan komponen kecenderungan berfrekuensi tinggi atau dengan kata lain data harga cabai harian memberikan kontribusi kecenderungan terbaik sehingga memberikan pengaruh lebih besar.
3. Total ragam yang mampu dijelaskan oleh semua IMF dan residu hanya 93,182% untuk data harga cabai harian dan 95,84% saja untuk data harga cabai mingguan.

## Saran

Penelitian selanjutnya dapat mengkaji kembali metode *Empirical Mode Decomposition* (EMD) dengan *ensemble*. Selain itu, banyak IMF yang dihasilkan dalam EMD penting untuk diperhatikan, terutama apabila terdapat banyak IMF dan residunya.

## DAFTAR PUSTAKA

Anonim,2010, Direktorat Jendral Pengolahan dan Pemasaran Hasil Pertanian, Kementerian Pertanian, Pedoman Umum Pelaksanaan Pengembangan Hortikultur. Jakarta.diakses tanggal 12 Maret 2016

<http://hortikultura.deptan.go.id>.

Box, G.E.P., Jenkins, G.M and Reinsel, G.C., 1994, *Time Series Analysis Forecasting and Control*, Third Edition. Prentice-Hall, Inc. USA.

Brooks. C., 2008, *Introductory Econometrics for Finance 2<sup>nd</sup> Edition*.Newyork : Cambridge University Press.

Warsito.B, dan Ispriyanti.D., 2004, *Uji Linieritas Data Time Series dengan RESET TEST*.Jurnal Matematika dan Komputer.

Cryer, J. D., 1986, *Time Series Analysis*, PWS-KENT Publishing Company, Boston.

Fransiska, H., 2014, Metode Dekomposisi Ensemble Untuk Memprediksi Harga Beras DKI Jakarta, *Thesis*, Bogor : IPB

Huang NE, Shen Z, Long SR, Wu MC, Shih HH, et al. 1998. *The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis*. Di dalam: Proceedings of the Royal Society London A454; London, 3 Juni 1996. hlm 903-995

Wei, W.W.S. 2006. *Time Series Analysis: Univariate and Multivariate Methods*. Addison-Wesley Publishing Company: New York.

# PENERAPAN MODEL LOG LINIER EMPAT DIMENSI PADA DATA JUMLAH DOSEN TETAP UNIVERSITAS BENGKULU

Fradelina Sari Tando<sup>1</sup>, Sigit Nugroho<sup>2</sup>, Etis Sunandi<sup>3</sup>

Mahasiswa Jurusan Matematika FMIPA Universitas Bengkulu, Bengkulu<sup>1</sup>

Email : [fradel.adel@gmail.com](mailto:fradel.adel@gmail.com)

Dosen Jurusan Matematika FMIPA Universitas Bengkulu, Bengkulu<sup>2,3</sup>

## ABSTRAK

Di dalam penelitian ini, peneliti membahas penerapan model log linier empat dimensi. Tujuan dari penulisan skripsi ini adalah untuk menjelaskan prosedur analisis menggunakan model log linier empat dimensi dan memodelkan model log linier dengan penerapan pada data jumlah dosen tetap Universitas Bengkulu. Hasil analisis data menggunakan data jumlah dosen tetap Universitas Bengkulu per 20 Maret 2017 bahwa model yang terpilih yaitu model (XY, WY, YZ, WX, WZ, XZ), model log liniernya yaitu :

$$\log m_{ijk} = \mu + \lambda_i^W + \lambda_j^X + \lambda_k^Y + \lambda_l^Z + \lambda_{ij}^{WX} + \lambda_{ik}^{WY} + \lambda_{il}^{WZ} + \lambda_{jk}^{XY} + \lambda_{jl}^{XZ} + \lambda_{kl}^{YZ}$$

Hal ini berarti bahwa faktor tingkat pendidikan (X) dan faktor usia (Y) saling berhubungan. Selain itu, jabatan akademik (W) dan usia (Y) saling berhubungan. Usia (Y) dan jenis kelamin (Z) saling berhubungan. Jabatan akademik (W) dan tingkat pendidikan (X) saling berhubungan. Jabatan akademik (W) dan jenis kelamin (Z) saling berhubungan. Tingkat pendidikan (X) dan jenis kelamin (Z) juga saling berhubungan.

**Kata kunci:** model, log linier, empat dimensi

## 1. PENDAHULUAN

Universitas Bengkulu (UNIB) merupakan satu-satunya Perguruan Tinggi Negeri di bawah naungan Kementerian Riset, Teknologi, dan Pendidikan Tinggi Republik Indonesia (disingkat Kemenristekdikti RI) yang ada di Provinsi Bengkulu. Universitas Bengkulu didirikan pada tahun 1982 dengan visi "UNIB menjadi universitas kelas dunia pada tahun 2025" (Anonim, 2012). Berdasarkan visi tersebut, maka Universitas Bengkulu harus memanfaatkan sumber daya manusia secara optimal dalam upaya mencapai misi dan tujuannya. Salah satu sumber daya manusia di Universitas Bengkulu adalah tenaga pengajar tetap (dosen tetap), sebagaimana terdapat dalam Peraturan Pemerintah Republik Indonesia Nomor 37 Tahun 2009 tentang Dosen, Bab I Ketentuan Umum Pasal 1 ayat (2), Dosen tetap adalah dosen yang bekerja penuh waktu yang berstatus sebagai tenaga pendidik tetap pada satuan pendidikan tinggi tertentu.

Berdasarkan data yang diperoleh dari Biro Umum dan Kepegawaian Universitas Bengkulu bagian Kepegawaian subbagian Tenaga Pendidik, jumlah dosen tetap Universitas Bengkulu per 20 Maret 2017 sebanyak 732 orang. Data jumlah dosen tetap

Universitas Bengkulu per 20 Maret 2017 ini berdasarkan jabatan akademik, tingkat pendidikan, usia, dan jenis kelamin. Data jumlah dosen tetap Universitas Bengkulu yang terkumpul dapat dikategorikan menjadi satu atau lebih kategori. Data seperti ini disebut data kategori. Salah satu metode untuk menganalisis data kategori adalah menggunakan model log linier. Model log linier digunakan untuk menganalisis hubungan antara variabel respon yang bersifat kategori yang membentuk tabel kontingensi (Agresti, 2002).

Penulis tertarik melakukan penelitian mengenai penerapan model log linier empat dimensi pada data jumlah dosen tetap Universitas Bengkulu berdasarkan jabatan akademik, tingkat pendidikan, usia, dan jenis kelamin. Tujuan dari penelitian ini adalah menjelaskan prosedur analisis menggunakan model log linier empat dimensi dan memodelkan model log linier empat dimensi pada data jumlah dosen tetap Universitas Bengkulu.

## 2. METODE PENELITIAN

### 2.1 Jenis Data

Jenis data dari penelitian ini adalah data sekunder. yaitu data populasi atau jumlah keseluruhan dosen tetap Universitas Bengkulu per 20 Maret 2017 sebanyak 732 orang berdasarkan jabatan akademik, tingkat pendidikan, usia, dan jenis kelamin dosen yang diperoleh dari Biro Umum dan Kepegawaian Universitas Bengkulu subbagian Tenaga Pendidik.

### 2.2 Variabel Penelitian

Variabel penelitian, definisi operasional, dan kategori dapat dilihat pada Tabel 1 di bawah ini :

**Tabel 1. Variabel penelitian, definisi operasional, dan kategori**

Variabel Penelitian	Definisi Operasional	Kategori
Jabatan akademik	Jabatan akademik dosen saat dilakukan penelitian	1 : tenaga pengajar 2 : asisten ahli 3 : lektor 4 : lektor kepala 5 : guru besar
Tingkat pendidikan	Pendidikan terakhir dosen saat dilakukan penelitian	1 : S1 2 : S1 Profesi 3 : S2 4 : S3
Usia	Usia dosen merupakan rentang kehidupan dalam tahun yang dihitung sejak dilahirkan sampai pada saat dilakukan penelitian	1 : 28-44 tahun 2 : 45-64 tahun 3 : 65-70 tahun
Jenis Kelamin	Jenis kelamin dosen	1 : laki-laki 2 : perempuan

### 2.3. Teknik Analisis Data

Langkah-langkah analisis yang digunakan pada penelitian ini adalah sebagai berikut:

1. Analisis deskriptif, dengan tujuan untuk menggambarkan kumpulan data atau hasil pengamatan pada data jumlah tenaga pengajar tetap Universitas Bengkulu berdasarkan jabatan akademik, tingkat pendidikan, usia, dan jenis kelamin.
2. Analisis model log linier untuk mengukur hubungan (asosiasi) antara empat variabel yang diteliti menggunakan metode log linier dengan bantuan *Software* SPSS versi 23 dengan langkah:
  - a. Menyusun tabel kontingensi empat dimensi berdasarkan empat variabel yang diteliti.
  - b. Membentuk model log linier empat dimensi.
  - c. Melakukan uji kesesuaian model secara simultan dengan 2 cara yaitu:
    - Uji pengaruh interaksi K-faktor yang lebih tinggi, dimana dalam penelitian ini faktor yang paling tinggi untuk K=4
    - Uji parameter pengaruh interaksi K-faktor
  - d. Melakukan uji asosiasi parsial untuk menganalisis pola hubungan antar keempat variabel yang diteliti.
  - e. Pemilihan model log linier empat dimensi dengan menyeleksi model terlengkap hingga model paling sederhana dengan menggunakan metode *backward elimination*.
  - f. Melakukan pengujian kecocokan model dengan kriteria uji yaitu:

Jika  $G^2 > \chi^2_{(\alpha,df)}$  maka tolak  $H_0$  atau dapat dikatakan model tidak diterima atau tidak layak digunakan (model tidak sesuai dengan keadaan sebenarnya). Jika  $G^2 \leq \chi^2_{(\alpha,df)}$  maka terima  $H_0$  atau dapat dikatakan bahwa model diterima atau layak digunakan (model sesuai dengan keadaan sebenarnya).
  - g. Menarik kesimpulan.

## 3. HASIL DAN PEMBAHASAN

### 3.1. Analisis Deskriptif Data Jumlah Dosen Tetap Universitas Bengkulu

Data pada penelitian ini diperoleh dari Biro Umum dan Kepegawaian Universitas Bengkulu subbagian Tenaga Pendidik per 20 Maret 2017 sebanyak 732 orang. Empat variabel yang digunakan yaitu jabatan akademik, tingkat pendidikan, usia, dan jenis kelamin. Tabel distribusi frekuensi pada data jumlah dosen tetap Universitas Bengkulu adalah sebagai berikut:

**Tabel 2. Tabel distribusi frekuensi data dosen tetap Universitas Bengkulu**

No	Variabel	Kategori	Frekuensi	Persentase (%)
1	Jabatan Akademik	1. Tenaga Pengajar	39	5.33
		2. Asisten Ahli	65	8.88
		3. Lektor	221	30.19
		4. Lektor Kepala	369	50.41
		5. Guru Besar	38	5.19
	Jumlah			732
2	Tingkat Pendidikan	1. S1	10	1.37
		2. S1 Profesi	18	2.46
		3. S2	471	64.34
		4. S3	233	31.83
	Jumlah			732
3	Usia	1. Muda	266	36.34
		2. Tengah Baya	461	62.98
		3. Tua	5	0.68
	Jumlah			732
4	Jenis Kelamin	1. Laki-laki	453	61.89
		2. Perempuan	279	38.11
	Jumlah			732

Pada Tabel 2 di atas dapat dilihat bahwa jabatan akademik dosen terbanyak adalah lektor kepala, tingkat pendidikan terbanyak adalah S2, usia dosen terbanyak adalah tengah baya, dan jenis kelamin terbanyak adalah laki-laki.

### 3.2 Prosedur Dalam Analisis Model Log Linier Empat Dimensi

#### 3.2.1 Tabel Kontingensi Empat Dimensi

Tabel kontingensi empat dimensi disusun dari data jumlah dosen tetap Universitas Bengkulu berdasarkan empat variabel yang diamati. Jabatan akademik dikelompokkan kedalam lima jenis jabatan akademik yaitu tenaga pengajar, asisten ahli, lektor, lektor kepala, dan guru besar. Selain itu, dosen tetap Universitas Bengkulu juga dikelompokkan menurut tingkat pendidikan, usia dan jenis kelamin.

#### 3.2.2 Pembentukan Model Log Linier Empat Dimensi pada Data Jumlah Dosen Tetap Universitas Bengkulu

Pembentukan model log linier empat dimensi pada data jumlah dosen tetap Universitas Bengkulu sesuai peubah amatan, dengan W sebagai variabel jabatan akademik, X sebagai variabel tingkat pendidikan, Y sebagai variabel usia, dan Z sebagai variabel jenis kelamin dinyatakan dalam persamaan (1) berikut ini:

$$\log m_{ijkl} = \mu + \lambda_i^W + \lambda_j^X + \lambda_k^Y + \lambda_l^Z + \lambda_{ij}^{WX} + \lambda_{ik}^{WY} + \lambda_{il}^{WZ} + \lambda_{jk}^{XY} + \lambda_{jl}^{XZ} + \lambda_{kl}^{YZ} + \lambda_{ijl}^{WXZ} + \lambda_{ikl}^{WYZ} + \lambda_{jkl}^{XYZ} + \lambda_{ijkl}^{WXYZ} \quad (1)$$

### 3.2.3 Uji Kesesuaian Model Secara Simultan

Uji kesesuaian model secara simultan untuk interaksi dari variabel jabatan akademik, tingkat pendidikan, usia, dan jenis kelamin dalam pengujian ini terdapat dua langkah pengujian pemodelan yaitu :

1. Uji untuk interaksi  $K$ -faktor dan yang lebih tinggi sama dengan nol.

Pengujian parameter dapat dilakukan yang bertujuan untuk menentukan parameter mana yang memberikan pengaruh di dalam model. Uji ini juga bertujuan untuk mengetahui jumlah variabel (faktor) tertinggi yang berasosiasi didalam model log linier, dimana parameter yang berpengaruh diuji dengan membandingkan rasio kemungkinan (*likelihood Rasio*) yaitu statistik  $G^2$  masing-masing model dengan  $\chi^2_{(\alpha;df)}$  pada tingkat kepercayaan tertentu.

**Tabel 3. Pengujian interaksi K-faktor dan yang lebih tinggi**

	K	df	Likelihood Ratio		Pearson		Jumlah Iterasi
			Chi-Square	Sig.	Chi-Square	Sig.	
Interaksi k-faktor dan yang lebih tinggi	1	119	2830.733	0.000	6661.770	0.000	0
	2	109	841.404	0.000	1891.078	0.000	2
	3	74	41.564	0.999	43.884	0.998	7
	4	24	1.000	1.000	0.509	1.000	10

Berdasarkan hasil yang diperoleh, untuk  $K = 4$  dan  $K = 3$ , memberi keputusan bahwa  $H_0$  diterima yang berarti bahwa interaksi 4 faktor dan interaksi 3 faktor tidak ada dalam model. Sedangkan untuk  $K = 2$  dan  $K = 1$ , memberi keputusan bahwa  $H_0$  ditolak, yang berarti secara signifikan menjelaskan model log linier yang akan di pakai sekurang-kurangnya harus memuat asosiasi 2 faktor.

2. Uji untuk interaksi  $K$ -faktor sama dengan nol.

Tahap pertama dimulai dengan pengujian orde yang lebih rendah yaitu  $K = 1$  sampai  $K = 4$ .

**Tabel 4. Pengujian interaksi K-faktor**

	K	Df	Likelihood Ratio		Pearson		Jumlah Iterasi
			Chi-Square	Sig.	Chi-Square	Sig.	
Interaksi k-faktor	1	10	1989.329	0.000	4770.692	0.000	0
	2	35	799.840	0.000	1847.194	0.000	0
	3	50	40.564	0.827	43.375	0.735	0
	4	24	1.000	1.000	0.509	1.000	0

Pengujian pengaruh asosiasi yang dimulai dari orde lebih rendah menghasilkan kesimpulan yang sama dengan pengujian dari orde lebih tinggi pada Tabel 4, bahwa model log linier yang dipakai sekurang-kurangnya harus memuat sebuah asosiasi dua faktor.

Berdasarkan hasil uji kesesuaian model secara simultan yang telah dilakukan, pengujian pengaruh asosiasi yang dimulai dari orde lebih rendah menghasilkan kesimpulan yang sama dengan pengujian dari orde lebih tinggi sehingga dapat diambil kesimpulan bahwa model log linier yang dapat dipakai sekurang-kurangnya harus memuat sebuah asosiasi dua faktor. Asumsi model lengkap untuk asosiasi dua faktor dapat dilihat pada persamaan (2) berikut ini:

$$\log m_{ijkl} = \mu + \lambda_i^W + \lambda_j^X + \lambda_k^Y + \lambda_l^Z + \lambda_{ij}^{WX} + \lambda_{ik}^{WY} + \lambda_{il}^{WZ} + \lambda_{jk}^{XY} + \lambda_{jl}^{XZ} + \lambda_{kl}^{YZ} \quad (2)$$

Pengujian dilanjutkan menggunakan uji asosiasi parsial untuk menentukan asosiasi peubah yang signifikan.

### 3.2.4 Uji Asosiasi Parsial

Tahap selanjutnya adalah menguji kebebasan secara parsial. Uji ini akan menunjukkan interaksi-interaksi yang ada dalam model. Berikut ini dilakukan uji parameter secara parsial dengan  $\alpha = 0.05$  dengan hipotesis sebagai berikut:

$H_0$  : Tidak ada interaksi antar variabel.

$H_1$  : Ada interaksi antar variabel

**Tabel 5. Uji asosiasi parsial**

Interaksi	df	Uji Parsial	Sig.	Jumlah Iterasi
W*X*Y	24	15.825	0.894	11
W*X*Z	12	25.519	0.013	4
W*Y*Z	8	7.158	0.520	20
X*Y*Z	6	4.123	0.660	10
W*X	12	179.039	0.000	7
W*Y	8	297.252	0.000	5
X*Y	6	15.832	0.015	7
W*Z	4	14.408	0.006	5
X*Z	3	8.212	0.042	7
Y*Z	2	49.608	0.000	5
W	4	553.035	0.000	2
X	3	861.478	0.000	2
Y	2	531.095	0.000	2
Z	1	43.721	0.000	2

Berdasarkan pengujian secara parsial, dapat disimpulkan bahwa terdapat interaksi antara jabatan akademik, tingkat pendidikan, dan jenis kelamin (WXZ), jabatan akademik dan tingkat pendidikan (WX), jabatan akademik dan usia (WY), tingkat pendidikan dan usia (XY), jabatan akademik dan jenis kelamin (WZ), tingkat pendidikan dan jenis kelamin (XZ), usia dan jenis kelamin (YZ) di dalam model. Pada model juga terdapat variabel utama jabatan akademik (W), tingkat pendidikan (X), usia (Y), dan jenis kelamin (Z).

### 3.2.5 Pemilihan Model Log Linier Empat Dimensi

Dari pengujian parameter pada langkah sebelumnya, diperoleh bahwa model mengandung asosiasi dua faktor. Metode yang digunakan untuk pemilihan model terbaik pada penelitian ini adalah metode *Backward Elimination*. Metode ini dimulai dengan menyeleksi model terlengkap. Setelah penyeleksian akan diperoleh model terbaik. Pemilihan model terbaik untuk melihat asosiasi jabatan akademik, tingkat pendidikan, usia, dan jenis kelamin pada data dosen tetap Universitas Bengkulu sesuai dengan Lampiran 1.

Dari pemilihan model di atas menunjukkan bahwa terdapat pengaruh interaksi variabel jabatan akademik dan tingkat pendidikan, variabel jabatan akademik dan usia, variabel jabatan akademik dan jenis kelamin, variabel tingkat pendidikan dan usia, variabel tingkat pendidikan dan jenis kelamin, serta variabel usia dan jenis kelamin dalam model sehingga pengaruh variabel utama jabatan akademik, variabel utama tingkat pendidikan, variabel utama usia, dan variabel utama jenis kelamin juga ada dalam model.

### 3.2.6 Pengujian Kecocokan Model

Pengujian kecocokan model pada model terbaik yang diperoleh dari metode *Backward Elimination* dilakukan dengan cara memeriksa hasil prediksi model dengan hasil pengamatan yang sebenarnya, dengan hipotesis sebagai berikut:

$H_0$  : Frekuensi harapan sesuai dengan model yang diasumsikan

$H_1$  : Frekuensi harapan tidak sesuai dengan model yang diasumsikan

Berdasarkan uji statistik kesamaan (*Likelihood Ratio*) Chi-kuadrat diperoleh nilai  $G^2 = 41.574$  dan  $p - value = 0.999$ . Dengan menggunakan taraf nyata = 0.05, maka  $p - value > \alpha$  sehingga dapat dikatakan model layak digunakan.

## 4. KESIMPULAN

Berdasarkan hasil dan pembahasan, maka dapat diambil kesimpulan bahwa penerapan model log linier dalam skripsi ini adalah menganalisis variabel-variabel yang saling berhubungan antara keempat variabel yang diteliti. Hasil analisis data menggunakan data jumlah dosen tetap Universitas Bengkulu per 20 Maret 2017 bahwa model yang terpilih yaitu model (XY, WY, YZ, WX, WZ, XZ), model log liniernya yaitu :

$$\log m_{ijkl} = \mu + \lambda_i^W + \lambda_j^X + \lambda_k^Y + \lambda_l^Z + \lambda_{ij}^{WX} + \lambda_{ik}^{WY} + \lambda_{il}^{WZ} + \lambda_{jk}^{XY} + \lambda_{jl}^{XZ} + \lambda_{kl}^{YZ}$$

Hal ini berarti bahwa faktor tingkat pendidikan (X) dan faktor usia (Y) saling berhubungan. Selain itu, jabatan akademik (W) dan usia (Y) saling berhubungan. Usia (Y) dan jenis kelamin (Z) saling berhubungan. Jabatan akademik (W) dan tingkat pendidikan

(X) saling berhubungan. Jabatan akademik (W) dan jenis kelamin (Z) saling berhubungan. Tingkat pendidikan (X) dan jenis kelamin (Z) juga saling berhubungan.

## 5. DAFTAR PUSTAKA

- [1] Anonim. 1999. *Keputusan Bersama Menteri Pendidikan dan Kebudayaan dan Kepala Badan Kepegawaian Negara, Nomor 181 Tahun 1999 Tentang Petunjuk Pelaksanaan Jabatan Fungsional Dosen dan Angka Kreditnya*. Jakarta: Kemendiknas. Diakses pada tanggal 3 Mei 2017. <http://sumberdaya.ristekdikti.go.id/wp-content/uploads/2016/02/Keputusan-bersama-Menteri-Pendidikan-dengan-KaBKN-tentang-Jafung-Dosen.pdf>
- [2] Anonim. 2005. *Undang-Undang Republik Indonesia Nomor 14 Tahun 2005 Tentang Guru dan Dosen*. Jakarta: Sekretariat Negara. Diakses pada tanggal 3 Mei 2017. [multisite.itb.ac.id/sa/wp-content/uploads/sites/44/2016/03/UU\\_14\\_2005.pdf](http://multisite.itb.ac.id/sa/wp-content/uploads/sites/44/2016/03/UU_14_2005.pdf)
- [3] Anonim. 2009. *Peraturan Pemerintah Republik Indonesia Nomor 37 Tahun 2009 Tentang Dosen*. Jakarta: Sekretariat Negara. Diakses pada tanggal 3 Mei 2017. [sipma.ui.ac.id/files/dokumen/...DOSEN/PP%2037%20Tahun%202009%20DOSEN.pdf](http://sipma.ui.ac.id/files/dokumen/...DOSEN/PP%2037%20Tahun%202009%20DOSEN.pdf)
- [4] Anonim. 2012. *Sejarah Universitas Bengkulu*. Diakses pada tanggal 27 Februari 2017. <http://www.unib.ac.id/profil/sejarah/>
- [5] Anonim. 2012. *Visi dan Misi Universitas Bengkulu*. Diakses pada tanggal 27 Februari 2017. <http://www.unib.ac.id/profil/819-2/>
- [6] Agresti, A. 2002. *Categorical Data Analysis Second Edition*. New York: John Wiley & Sons, Inc.

Lampiran 1. Pemilihan Model Log Linier

Langkah		Interaksi	Chi-Square	df	Sig.	Jumlah Iterasi	
0	Model Umum	$W^*X^*Y^*Z$	0.000	0	.		
	Hapus Interaksi 1	$W^*X^*Y^*Z$	1.000	24	1.000	10	
1	Model Umum	$W^*X^*Y, W^*X^*Z, W^*Y^*Z, X^*Y^*Z$	1.000	24	1.000		
	Hapus Interaksi	1	$W^*X^*Y$	15.825	24	0.894	11
		2	$W^*X^*Z$	25.519	12	0.013	4
		3	$W^*Y^*Z$	7.158	8	0.520	20
		4	$X^*Y^*Z$	4.123	6	0.660	10
2	Model Umum	$W^*X^*Z, W^*Y^*Z, X^*Y^*Z$	16.824	48	1.000		
	Hapus Interaksi	1	$W^*X^*Z$	21.139	12	0.048	6
		2	$W^*Y^*Z$	3.664	8	0.886	20
		3	$X^*Y^*Z$	0.599	6	0.996	11
3	Model Umum	$W^*X^*Z, W^*Y^*Z, X^*Y$	17.423	54	1.000		
	Hapus Interaksi	1	$W^*X^*Z$	21.106	12	0.049	6
		2	$W^*Y^*Z$	3.915	8	0.865	20
		3	$X^*Y$	26.632	6	0.000	2
4	Model Umum	$W^*X^*Z, X^*Y, W^*Y, Y^*Z$	21.338	62	1.000		
	Hapus Interaksi	1	$W^*X^*Z$	20.235	12	0.063	9
		2	$X^*Y$	25.974	6	0.000	7
		3	$W^*Y$	307.376	8	0.000	4
		4	$Y^*Z$	59.741	2	0.000	6
5	Model Umum	$X^*Y, W^*Y, Y^*Z, W^*X, W^*Z, X^*Z$	41.574	74	0.999		
	Hapus Interaksi	1	$X^*Y$	15.835	6	0.015	7
		2	$W^*Y$	297.242	8	0.000	4
		3	$Y^*Z$	49.598	2	0.000	6
		4	$W^*X$	179.029	12	0.000	9
		5	$W^*Z$	14.400	4	0.006	6
		6	$X^*Z$	8.201	3	0.042	10
6	Model Umum	$X^*Y, W^*Y, Y^*Z, W^*X, W^*Z, X^*Z$	41.574	74	0.999		

# **PENGELOMPOKAN PROVINSI DI INDONESIA BERDASARKAN RATA-RATA PRODUKSI TANAMAN PANGAN MENGGUNAKAN METODE KLASTER K-MEANS**

**Dian Juli Adisaputra<sup>1</sup> Sigit Nugroho<sup>2</sup> Pepi Novianti<sup>2</sup>**

<sup>1</sup> Alumni Matematika FMIPA Universitas Bengkulu

<sup>2</sup> Staf Pengajar Matematika FMIPA Universitas Bengkulu

## **ABSTRACT**

This research applying k-means cluster analysis into classification of provinces in Indonesia based on the average production of food crops. The results of this study indicate that members of the Province in cluster 4 (Riau, Jambi, Bengkulu, Bangka Belitung, Riau Islands, Yogyakarta, Bali, East Nusa Tenggara, West Kalimantan, Central Kalimantan, East Kalimantan, North Kalimantan, North Sulawesi, Central Sulawesi, Southeast Sulawesi, Gorontalo, West Sulawesi, Maluku, North Maluku, West Papua and Papua) is a region with low average production of food crops. While members of the Province in cluster 2 (East Java) is a region with higher average production of food crops.

Keywords: Cluster Analysis, Non Hierarchical Cluster, K-Means, food crops

## **PENDAHULUAN**

Pertanian merupakan salah satu sektor utama dalam pembangunan ekonomi di Indonesia, mengingat Indonesia merupakan daerah agraris. Selain kontribusinya dalam Produk Domestik Bruto (PDB), peranan sektor pertanian dalam pembangunan ekonomi sangat luas, diantaranya sebagai sektor penyerap tenaga kerja terbesar, sebagai penghasil makanan penduduk, dan sebagai penentu stabilitas harga. Pemerintah juga mengharapkan produksi pertanian tanaman pangan mengalami peningkatan tiap tahunnya (Rivani, 2010).

Tanaman Palawija merupakan tanaman yang potensial untuk dikembangkan karena hasilnya dapat digunakan sebagai sumber karbohidrat, sumber protein nabati, dan bahan dasar berbagai industri. Tanaman palawija meliputi jagung, kedelai, kacang tanah, kacang hijau, ubi kayu, dan ubi jalar (Safitri, 2012).

Dalam Undang-Undang Republik Indonesia Nomor 7 Tahun 1996, dikenal dua istilah penting tentang pangan, yaitu sistem pangan dan ketahanan pangan. Sistem pangan diartikan sebagai segala sesuatu yang berhubungan dengan peraturan, pembinaan, dan pengawasan terhadap kegiatan atau produksi pangan dan peredaran pangan sampai dengan siap dikonsumsi oleh manusia. Sementara itu, ketahanan pangan diartikan sebagai kondisi terpenuhinya pangan bagi rumah tangga yang tercermin dari tersedianya pangan yang cukup, baik jumlah maupun mutunya, aman, merata, dan terjangkau. Pangan diartikan sebagai segala sesuatu yang bersumber dari sumber hayati dan air, baik yang diolah maupun yang tidak diolah. Pangan diperuntukkan bagi konsumsi manusia sebagai makanan atau minuman, termasuk bahan tambahan pangan, bahan baku pangan, dan bahan-bahan lain yang digunakan dalam proses penyiapan, pengolahan, dan pembuatan makanan atau minuman (Anonim, 2014).

Analisis kluster pertama kali digunakan oleh Tyron pada tahun 1939. Analisis kluster bertujuan untuk mengalokasikan sekelompok individu pada suatu kelompok-kelompok yang saling bebas sehingga individu-individu di dalam satu

kelompok yang sama mirip satu sama lain, sedangkan individu-individu di dalam kelompok yang berbeda tidak mirip (Rachmatin, 2014).

Terdapat beberapa metode pengelompokan dalam analisis kluster yaitu analisis kluster secara hierarki dan non hierarki, di dalam analisis kluster non hierarki penulis akan menerapkan metode kluster *K-Means*, metode ini dimulai dengan menentukan terlebih dahulu jumlah kluster yang diinginkan dan *centroid* di tiap kluster (Nugroho, 2008). *K-Means* merupakan salah satu metode kluster non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih kelompok, sehingga data yang memiliki karakteristik sama dikelompokkan ke dalam satu kelompok yang sama (Rismawan dan Kusumadewi, 2008).

Beberapa peneliti yang lebih dahulu membahas penelitian yang berkaitan dengan analisis Kluster *K-Means* diantaranya yaitu : Susanti (2012) melakukan penelitian dengan judul Pengklasteran Provinsi di Indonesia Berdasarkan Jumlah Penderita Penyakit Penyebab Kematian Tahun 2012 Dengan Menggunakan Metode *K-Means*. Safitri (2012) melakukan penelitian dengan judul Analisis *Cluster* Pada Kabupaten/Kota Di Jawa Tengah Berdasarkan Produksi Palawija. Rismawan dan Kusumadewi (2008) melakukan penelitian dengan judul Aplikasi *K-Means* Untuk Pengelompokkan Mahasiswa Berdasarkan Nilai *Body Mass Index* (BMI) & Ukuran Kerangka.

Berdasarkan penjelasan sebelumnya penulis tertarik untuk melakukan penelitian dengan judul : Pengelompokan Provinsi Di Indonesia Berdasarkan Rata-Rata Produksi Tanaman Pangan Menggunakan Metode Kluster *K-Means*.

## **TINJAUAN PUSTAKA**

### **2.1 Tanaman Pangan**

Tanaman pangan adalah segala jenis tanaman yang dapat menghasilkan karbohidrat dan protein. Tanaman pangan terdiri dari padi, jagung, kedelai, kacang tanah, kacang hijau, ubi kayu, dan ubi jalar. Indonesia adalah negara agraris tempat tumbuh jenis tanaman pangan, dahulu indonesia pernah dikenal sebagai negara swasembada pangan (Hanum, 2008).

### **2.2 Analisis Kluster**

Analisis kluster merupakan analisis yang digunakan untuk mengelompokkan pengamatan atau variabel menjadi beberapa kelompok pengamatan atau variabel yang jumlahnya lebih sedikit. Analisis kluster bertujuan untuk mengelompokkan  $n$  objek berdasarkan  $p$  variat yang memiliki kesamaan karakteristik diantara objek-objek tersebut. Objek tersebut akan diklasifikasikan ke dalam satu atau lebih kluster sehingga objek-objek yang berada dalam satu kluster akan mempunyai kemiripan atau kesamaan karakter (Yulianto, 2014).

### **2.3 Metode Kluster Non Hierarki**

Metode ini dimulai dengan menentukan terlebih dahulu jumlah kluster yang diinginkan (dua kluster atau lebih). Setelah jumlah kluster ditentukan, baru proses kluster dilakukan tanpa mengikuti proses hierarki. Kelemahan pengklasteran dengan menggunakan metode ini adalah banyaknya kluster harus ditentukan sebelumnya dan pemilihan pusat kluster secara sembarang. Keunggulan pengklasteran dengan menggunakan metode ini lebih cepat daripada dengan pengklasteran dengan menggunakan metode hierarki dan lebih menguntungkan untuk jumlah objek/kasus yang besar (Rosiatun, 2010).

### **2.4 Proses Analisis Kluster**

Langkah-langkah yang harus dilakukan untuk melakukan proses analisis kluster yaitu (Santoso, 2002) :

1. Mengukur kesamaan antar objek (*similarity*).
  1. Mengukur korelasi antar sepasang objek pada beberapa variabel.
  2. Mengukur jarak antara dua objek, pengukuran ada bermacam-macam dan yang paling populer adalah menggunakan metode *Euclidean Distance*.

2. Membuat kluster (dengan metode secara hierarki atau secara non hierarki).
3. Setelah kluster terbentuk, lakukan interpretasi terhadap kluster yang telah terbentuk, untuk menggambarkan isi kluster tersebut.
4. Melakukan validasi dan *profiling*.  
Kluster yang terbentuk kemudian diuji apakah hasil pengklasteran tersebut valid, kemudian dilakukan proses *profiling* untuk menjelaskan karakteristik setiap kluster berdasarkan profil tertentu.

## 2.5 Kluster K-Means

Penggunaan pengklasteran *K-Means* untuk menjelaskan algoritma dalam

penentuan suatu obyek ke dalam kluster tertentu berdasarkan rataan terdekat. Asumsikan  $n$  adalah obyek dan  $p$  adalah variabel yang dinotasikan dengan  $D(i, j)$  dimana  $i = 1, 2, \dots, n$  dan  $j = 1, 2, \dots, p$  dan dengan menggunakan jarak *euclid* antar obyek. Pengklasteran *K-Means* sangat cocok untuk data dengan ukuran yang besar karena memiliki kecepatan yang lebih tinggi dibandingkan metode hierarki. Namun, pemilihan banyaknya kluster dan *centroid* yang harus ditentukan lebih dahulu menjadi kelemahan metode ini (Nugroho, 2008).

*K-Means* merupakan metode kluster non hierarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih kelompok. Data-data yang memiliki karakteristik yang sama dikelompokkan dalam satu kluster yang sama dan data yang memiliki karakteristik yang berbeda dikelompokkan dengan kluster yang lain (Agusta, 2007).

## 2.6 Algoritma K-Means Kluster

Algoritma *K-Means* merupakan algoritma yang relatif sederhana untuk mengklasifikasikan atau mengelompokkan sejumlah besar obyek dengan atribut tertentu ke dalam kelompok-kelompok sebanyak  $K$ . Pada algoritma *K-Means*, jumlah kluster  $K$  sudah ditentukan lebih dahulu. Data clustering menggunakan metode *K-Means* ini secara umum dilakukan dengan algoritma dasar sebagai berikut (Agusta, 2007) :

1. Tentukan jumlah kluster  $k$
2. Alokasikan data ke dalam kluster secara random
3. Hitung jarak setiap data yang ada terhadap setiap pusat kluster di masing-masing kluster menggunakan rumus *Euclidian Distance* (Nugroho, 2008) :

$$D(x_i, y_j) = \sqrt{(x_{i_1} - y_{j_1})^2 + (x_{i_2} - y_{j_2})^2 + \dots + (x_{i_p} - y_{j_p})^2} \quad (2.1)$$

$$D(x_i, y_j) = \sqrt{\sum_{l=1}^p (x_{i_l} - y_{j_l})^2} \quad (2.2)$$

Keterangan :

$D(x_i, y_j)$  = Jarak data ke- $i$  ke pusat kluster ke- $j$   
 $x_{i_l}$  = Data ke- $i$  pada variabel ke- $l$   
 $y_{j_l}$  = Titik pusat kluster ke- $j$  pada variabel ke- $l$

4. Kelompokkan setiap data ke jarak terdekat pusatnya
5. Tentukan posisi pusat kluster baru  $C_{kj}$  dengan cara menghitung nilai rata-rata dari data-data yang ada pada kluster yang sama dengan rumus (Rivani, 2010) :

$$C_{kj} = \left(\frac{1}{n_k}\right) \sum_{l=1}^p d_{j_l} \quad (2.3)$$

Keterangan :

$C_{kj}$  = Pusat kluster baru ke- $j$  pada variabel ke- $l$   
 $n_k$  = Banyaknya anggota objek dalam kluster ke- $k$   
 $d_{j_l}$  = Data dalam kluster ke- $k$  pada variabel ke- $l$

5. Jika pusat klaster tidak berubah lagi maka proses klaster selesai, atau kembali ke langkah 3 apabila masih ada data yang berpindah klaster.

### **2.7 Kelebihan dan Kekurangan *K-Means***

Kelebihan *K-Means* :

1. Algoritma *K-Means* ini relatif sederhana dan cepat.
2. Algoritma *K-Means* memberikan hasil yang relatif baik pada klaster.
3. Algoritma *K-Means* memiliki ketelitian yang cukup tinggi terhadap ukuran objek dan tidak berpengaruh pada urutan objek.

Kekurangan *K-Means* :

1. Menentukan jumlah klaster sebelum mengetahui jumlah klaster.
2. Hanya bekerja pada data numerik.
3. Sangat sensitif pada pembangkitan titik pusat awal secara random.

## **METODE PENELITIAN**

### **3.1 Jenis Penelitian**

Jenis penelitian ini adalah penelitian statistik terapan, penelitian ini dilakukan dengan menggunakan analisis Klaster *K-Means*.

### **3.2 Pengumpulan Data**

Data yang digunakan pada penelitian ini adalah data sekunder, yang diperoleh dari Badan Pusat Statistik (BPS) yaitu data produksi tanaman pangan diseluruh Provinsi di Indonesia pada tahun 2015.

### **3.3 Variabel Penelitian**

Objek yang digunakan dalam penelitian ini adalah Provinsi diseluruh Indonesia, dengan indikator penelitian adalah produksi tanaman pangan, dan variabel yang digunakan adalah sebagai berikut:

- X1 = Produksi Padi diseluruh Provinsi di Indonesia
- X2 = Produksi Jagung diseluruh Provinsi di Indonesia
- X3 = Produksi Kedelai diseluruh Provinsi di Indonesia
- X4 = Produksi Kacang Tanah diseluruh Provinsi di Indonesia
- X5 = Produksi Ubi Jalar diseluruh Provinsi di Indonesia

### **3.4 Metode Analisis**

Metode analisis yang digunakan dalam penelitian adalah analisis Klaster *K-Means*, dengan mengambil data hasil produksi tanaman pangan (padi, jagung, kedelai, kacang tanah, ubi jalar) di Provinsi di Indonesia pada tahun 2015 yang diperoleh dari Badan Pusat Statistik. Pengolahan data dilakukan dengan bantuan *software Microsoft Office Excel 2007* dan *SPSS 16.0*.

### **3.5 Teknik Analisis Data**

Dalam penelitian ini dianalisis melalui tahapan-tahapan sebagai berikut :

1. Menentukan jumlah klaster yang ingin dibentuk dengan metode *K-Means*
2. Memilih ukuran jarak antar objek yaitu dengan menggunakan jarak *Euclid*.
3. Melakukan proses pengklasteran, jika pada proses iterasi pengklasteran objek tidak ada yang berpindah maka proses iterasi pengklasteran selesai, jika masih ada objek yang berpindah maka lakukan proses iterasi pengklasteran.
4. Melakukan interpretasi pada klaster yang diperoleh.

## **BAB IV : HASIL DAN PEMBAHASAN**

### **4.1 Deskripsi Data**

Data yang digunakan dalam penelitian ini adalah data sekunder, yang diperoleh dari Badan Pusat Statistik (BPS) yaitu data produksi tanaman pangan diseluruh provinsi di Indonesia pada tahun 2015. Variabel data yang digunakan

dalam penelitian ini terdiri dari 5 variabel produksi tanaman pangan yaitu produksi padi, produksi jagung, produksi kedelai, produksi kacang tanah, dan produksi ubi jalar. Objek data yang digunakan dalam penelitian ini yaitu seluruh provinsi di Indonesia yang terdiri dari 33 provinsi.

#### 4.2 Metode Pengklasteran *K-Means*

Metode pengklasteran *K-Means* memproses semua data secara sekaligus, dimana *K* adalah banyaknya klaster. Dengan menggunakan bantuan program aplikasi SPSS 16.0 semua objek telah diproses dan diperoleh bahwa tidak ada data yang hilang seperti tabel *number of cases in each cluster* di bawah ini (Lampiran 2) sehingga proses pengklasteran dapat dilanjutkan.

**Tabel 1. Number of Cases in each Cluster**

Cluster	1	5.000
	2	1.000
	3	4.000
	4	22.000
	5	2.000
Valid		34.000
Missing		.000

#### 4.3 Pembentukan Klaster *K-Means*

Banyak klaster yang dibentuk peneliti pada proses pengklasteran dengan metode *K-Means* adalah sebanyak 5 klaster yaitu klaster 1, klaster 2, klaster 3, klaster 4, dan klaster 5. Dengan menggunakan bantuan program aplikasi SPSS 16.0, maka diperoleh :

**Tabel 2. Initial Cluster Centers**

	Cluster				
	1	2	3	4	5
Padi	2417392	13154967	5471806	959	11301422
Jagung	959973	6131163	1528414	473	3212391
Kedelai	125036	344998	67192	15	129794
KacangTanah	31142	191579	19024	153	109204
UbiJalar	19024	350516	71681	1795	151312

Tabel *Initial Cluster* di atas menunjukkan bahwa ada lima buah titik awal pusat klaster yang pertama kali terbentuk di setiap klaster, selanjutnya dengan metode klaster *K-Means* akan diuji dan dilakukan realokasi klaster yang ada (Lampiran 6, 7 dan 8). Proses tersebut dinamakan *Iteration*, yang memuat perubahan pada pusat klaster.

#### 4.4 Analisis Pengelompokan Berdasarkan Produksi Tanaman Pangan

Dari lima variabel yang relevan, untuk membedakan isi klaster, dapat dianalisis provinsi yang termasuk klaster 1, klaster 2, klaster 3, klaster 4, dan klaster 5. Pengelompokan 34 provinsi adalah sebagai berikut:

**Tabel 3. Hasil Pengelompokan Berdasarkan Produksi Tanaman Pangan**

<b>Klaster</b>	<b>Jumlah Anggota</b>	<b>Anggota</b>
1	5	Aceh
		Sumatera Barat
		Banten
		Nusa Tenggara Barat
		Kalimantan Selatan
2	1	Jawa Timur
3	4	Sumatera Utara
		Sumatera Selatan
		Lampung
		Sulawesi Selatan
4	21	Riau
		Jambi
		Bengkulu
		Kepulauan Bangka Belitung
		Kepulauan Riau
		Daerah Istimewa Yogyakarta
		Bali
		Nusa Tenggara Timur
		Kalimantan Barat
		Kalimantan Tengah
		Kalimantan Timur
		Kalimantan Utara
		Sulawesi Utara
		Sulawesi Tengah
		Sulawesi Tenggara
		Gorontalo
		Sulawesi Barat
		Maluku
		Maluku Utara
Papua Barat		
Papua		
5	2	Jawa Barat
		Jawa Tengah

Berdasarkan tabel di atas terlihat bahwa pengelompokan 33 provinsi ke dalam lima klaster produksi tanaman pangan adalah sebagai berikut: Klaster 1 terdiri dari Provinsi Aceh, Sumatera Barat, Banten, Nusa Tenggara Barat, dan Kalimantan Selatan. Klaster 2 hanya terdiri oleh Provinsi Jawa Timur. Klaster 3 terdiri dari Provinsi Sumatera Utara, Sumatera Selatan, Lampung, dan Sulawesi Selatan. Klaster 4 terdiri dari Provinsi Riau, Jambi, Bengkulu, Kepulauan Bangka Belitung, Kepulauan Riau, Daerah Istimewa Yogyakarta, Bali, Nusa Tenggara

Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat, dan Papua. Klaster 5 terdiri dari Provinsi Jawa Barat dan Jawa Tengah.

#### 4.5 Interpretasi Profil

Interpretasi profil klaster meliputi pengkajian mengenai *centroid* yaitu rata-rata nilai obyek yang terdapat dalam klaster pada tiap variabel (Nugroho, 2008). Interpretasi profil diperoleh dari menghitung nilai rata-rata obyek terhadap kelima klaster yang terbentuk dengan menganalisis variabel-variabel yang membedakan antara kelima klaster. Setelah diketahui variabel-variabel yang relevan untuk membedakan klaster 1, klaster 2, klaster 3, klaster 4, dan klaster 5, selanjutnya dianalisis rata-rata produksi masing-masing provinsi yang termasuk ke dalam masing-masing klaster yang dapat dilihat dari nilai akhir pusat klaster pada variabel yang membedakan.

**Tabel 4. Final Cluster Centers**

	Cluster				
	1	2	3	4	5
Padi	2325664	13154967	4351613	477130	11337328
Jagung	381604	6131163	1209851	116400	2086162
Kedelai	38225	344998	25094	4449	114366
KacangTanah	11952	191579	8631	5879	94962
UbiJalar	45389	350516	59838	38038	303744

Tabel *final cluster centers* diatas menunjukkan bahwa ada lima buah titik akhir pusat klaster yang terbentuk disetiap klaster. Karakteristik dalam setiap klaster ini berbeda, dapat dilihat dari nilai tabel *final cluster centers* di atas dari nilai satuan yang terkecil sampai nilai satuan yang terbesar disetiap variabel, maka berturut-turut adalah sebagai daerah produksi paling sedikit, sedikit, sedang, banyak, dan paling banyak untuk nilai satuan yang terbesar di setiap variabel.

Secara umum provinsi pada klaster 2 (Jawa Timur) memproduksi tanaman pangan dengan jumlah rata-rata produksi yang paling banyak dibandingkan dengan jumlah produksi tanaman pangan pada klaster lainnya

Provinsi pada klaster 5 (Jawa Barat dan Jawa Tengah) merupakan provinsi dengan jumlah rata-rata produksi tanaman pangan banyak untuk padi, jagung, kedelai, kacang tanah, dan ubi jalar. Provinsi pada klaster 3 (Sumatera Utara, Sumatera Selatan, Lampung, dan Sulawesi Selatan) merupakan provinsi dengan jumlah rata-rata produksi sedang untuk padi, jagung dan ubi jalar, sedangkan untuk jumlah rata-rata produksi kedelai dan kacang tanah sedikit. Provinsi pada klaster 1 (Aceh, Sumatera Barat, Banten, Nusa Tenggara Barat, dan Kalimantan Selatan) merupakan provinsi dengan jumlah rata-rata produksi sedikit untuk padi, jagung dan ubi jalar, sedangkan untuk jumlah rata-rata produksi kedelai dan kacang tanah sedang.

Provinsi pada klaster 4 (Riau, Jambi, Bengkulu, Kepulauan Bangka Belitung, Kepulauan Riau, Daerah Istimewa Yogyakarta, Bali, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat, dan Papua) merupakan

provinsi dengan jumlah rata-rata produksi tanaman pangan sangat untuk padi, jagung, kedelai, kacang tanah, dan ubi jalar

## **BAB V : KESIMPULAN DAN SARAN**

### **5.1 Kesimpulan**

Pengelompokan Provinsi di Indonesia menggunakan analisis kluster *K-Means* berdasarkan rata-rata produksi tanaman pangan dikelompokkan menjadi 5 kluster, secara umum Provinsi pada kluster 2 (Jawa Timur) memproduksi tanaman pangan dengan jumlah rata-rata produksi yang paling banyak jika dibandingkan dengan jumlah produksi tanaman pangan pada kluster lainnya, sedangkan Provinsi pada kluster 4 (Riau, Jambi, Bengkulu, Kepulauan Bangka Belitung, Kepulauan Riau, Daerah Istimewa Yogyakarta, Bali, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat, dan Papua) memproduksi tanaman pangan dengan jumlah rata-rata produksi yang paling sedikit.

### **5.2 Saran**

Pada penelitian ini penulis menggunakan metode kluster *K-Means* untuk mengelompokkan provinsi di Indonesia berdasarkan rata-rata produksi tanaman pangan. Bagi peneliti selanjutnya yang tertarik dengan analisis kluster *K-Means* dapat menggunakan data dalam bidang lain, dan dikarenakan cakupan kluster sebenarnya sangat luas, maka bagi peneliti selanjutnya juga dapat menggunakan perbandingan kluster non hierarki dengan metode *K-Means* dengan kluster hierarki dengan metode lain.

## **DAFTAR PUSTAKA**

- Agusta, Y. 2007. *K-Means – Penerapan, Permasalahan dan Metode Terkait. Jurnal Sistem dan Informatika*, Vol. 3 : 47-60.
- Anonim. 2014. Dinas Pertanian Tanaman Pangan dan Hortikultura Provinsi Kalimantan Selatan. <http://distantph.kalselprov.go.id/2014/02/03/mengenal-tanaman-pangan> diakses pada Sabtu, 06 Februari 2016.
- Hanum, C. 2008. *Teknik Budidaya Tanaman*, Jilid Dua. Penerbit Direktorat Pembinaan Sekolah Menengah Kejuruan. Jakarta.
- Martiana, E., N, R, Muhtada'i., dan E, Purnomo. 2011. Penggunaan Metode Pengklasteran Untuk Menentukan Bidang Tugas Akhir Mahasiswa Teknik Informatika. <http://repo.eepis-its.edu/130> diakses pada Sabtu, 06 Februari 2016.
- Nugroho, S. 2008. *Pengantar Statistika Matematika*, Edisi Pertama. UNIB Press.
- Nugroho, S. 2008. *Statistika Multivariat Terapan*, Edisi Pertama. UNIB Press.
- Rachmatin, D. 2014. Aplikasi Metode-Metode Agglomerative Dalam Analisis Kluster Pada Data Tingkat Polusi Udara. *Jurnal Ilmiah Program Studi Matematika STKIP Siliwangi Bandung*, Vol 3, No.2.
- Rismawan, T, dan Kusumadewi, S. 2008. Aplikasi K-Means Untuk Pengelompokkan Mahasiswa Berdasarkan Nilai Body Mass Index (BMI) & Ukuran Kerangka. *Jurnal Seminar Nasional Aplikasi Teknologi Informasi (SNATI)*, ISSN: 1907-5022.
- Rivani, E. 2010. Aplikasi K-Means Cluster Untuk Pengelompokkan Provinsi Berdasarkan Produksi Padi, Jagung, Kedelai, Dan Kacang Hijau Tahun 2009. *Jurnal Mat Stat*, Vol. 10 No. 2 : 122-134.
- Sa'adah, I, R., Supriyanta., dan Subejo. 2013. Keragaman Warna Gabah Dan Warna Beras Varietas Lokal Padi Beras Hitam (*Oryza Sativa L.*) Yang

- Dibudidayakan Oleh Petani Kabupaten Sleman, Bantul, Dan Magelang. *Jurnal Vegetalika*, Vol.2 No.3 : 13-20.
- Safitri, D. 2012. Analisis Cluster Pada Kabupaten/Kota Di Jawa Tengah Berdasarkan Produksi Palawija. *Jurnal Media Statistika*, Vol. 5, No. 1 : 11-16.
- Susanti, P. 2012. Pengklasteran Provinsi di Indonesia Berdasarkan Jumlah Penderita Penyakit Penyebab Kematian Tahun 2012 Dengan Menggunakan Metode K-Means. *Skripsi Universitas Bengkulu*.
- Yulianto, S. 2014. Analisis Klaster Untuk Pengelompokan Kabupaten/Kota Di Provinsi Jawa Tengah Berdasarkan Indikator Kesejahteraan Rakyat. *Jurnal Statistika*, Vol. 2, No. 1.

# PERBANDINGAN METODE REGRESI LOGISTIK BINER DAN METODE CART PADA FAKTOR-FAKTOR YANG MEMPENGARUHI PENYAKIT JANTUNG KORONER (Studi Kasus di RSUD Dr. M. Yunus Bengkulu)

Wina Ayu Lestari<sup>1</sup>, Sigit Nugroho<sup>2</sup>, Fachri Faisal<sup>2</sup>

<sup>1</sup>Alumni Jurusan Matematika Fakultas MIPA Universitas Bengkulu

<sup>2</sup>Staf Pengajar Jurusan Matematika Fakultas MIPA Universitas Bengkulu

Email: [wina.yhrd@gmail.com](mailto:wina.yhrd@gmail.com)

Abstrak— Penelitian ini bertujuan untuk mengetahui faktor-faktor yang mempengaruhi penyakit jantung koroner. Analisis yang dapat digunakan untuk membuat klasifikasi adalah metode regresi logistik biner dan metode CART (*Classification and regression trees*). Data yang digunakan pada penelitian ini berupa data sekunder yang diperoleh dari data rekam medis di Rumah Sakit M. Yunus Bengkulu pada Tahun 2014 dengan variabel respon berskala katagorik. Pengklasifikasian menggunakan regresi logistik biner menghasilkan ketepatan klasifikasi sebesar 84.2% dimana terdapat empat faktor yang berpengaruh yaitu dislipidemia, hipertensi, diabetes melitus dan obesitas. Sedangkan dengan menggunakan metode CART terdapat tiga faktor yang mempengaruhi yaitu dislipidemia, diabetes melitus dan obesitas dengan ketepatan klasifikasi sebesar 95 %.

Kata Kunci : *PJK, Klasifikasi, Regresi Logistik Biner, CART*

## I. PENDAHULUAN

Penyakit jantung (kardiovaskuler) merupakan penyakit yang sangat menakutkan dan mematikan. Jumlah penderita penyakit jantung di seluruh dunia selalu bertambah setiap tahunnya. Penyakit jantung koroner (PJK) merupakan jenis gangguan pada jantung yang paling sering ditemui dan merupakan penyebab utama kematian di negara maju dan berkembang termasuk Indonesia (Zakiyah dkk., 2012). Menurut Soeharto (2004), penyakit jantung koroner adalah kelainan pada jantung yang disebabkan oleh terjadinya penyempitan atau penyumbatan arteri yang mengalirkan darah ke otot jantung.

Melihat permasalahan tersebut, penulis tertarik untuk mengetahui hubungan antara penyakit jantung koroner dan faktor-faktor risikonya dengan klasifikasi. Pada penelitian ini penulis membatasi faktor – faktor risiko yang akan diteliti yaitu umur, jenis kelamin, dislipidemia, tekanan darah tinggi (*hipertensi*), penyakit diabetes melitus dan obesitas. Studi kasus pada penelitian ini adalah di RSUD Dr. M. Yunus Bengkulu. Hal ini dikarenakan RSUD Dr. M. Yunus merupakan salah satu rujukan tempat berobat di Bengkulu.

Analisis yang dapat digunakan untuk membuat klasifikasi adalah metode regresi logistik biner dan CART (*Classification and regression trees*). Kedua analisis tersebut akan menghasilkan ketepatan klasifikasi tersendiri dan selanjutnya akan dilihat analisis mana yang memberikan ketepatan klasifikasi tertinggi (Margasari, 2014).

## II. TINJAUAN PUSTAKA

### Multikolinieritas

Multikolinieritas adalah suatu kondisi dimana terdapat dua atau lebih variabel prediktor yang digunakan dalam regresi saling berkorelasi. Salah satu cara untuk mendeteksi multikolinieritas adalah dengan melihat besarnya nilai *Variance Inflation Factor* (VIF) untuk masing-masing variabel prediktor

$$VIF_j = \frac{1}{1-R_j^2}$$

dimana :

$R_j^2$  : Koefisien determinasi berganda dari variabel prediktor  $X_j$  dengan seluruh variabel prediktor  $X$  lainnya.

Jika nilai  $VIF_j$  lebih dari atau sama dengan 10, maka ini menunjukkan data mengalami masalah multikolinieritas (Neter

dkk., 1990). Salah satu cara untuk mengatasi multikolinieritas adalah dengan mengeluarkan satu atau beberapa variabel prediktor yang terindikasi memiliki hubungan dengan variabel prediktor lainnya (Gujarati, 2010).

### Analisis Regresi Logistik Biner

Regresi logistik biner merupakan salah satu metode regresi logistik yang digunakan untuk mencari hubungan antara variabel respon ( $y$ ) yang bersifat *biner* atau dikotomis dengan variabel prediktor ( $x$ ) yang bersifat dikotomis atau polikotomis (Hosmer dan Lemeshow, 2000). Variabel respon ( $y$ ) terdiri dari 2 kategori yaitu 0 dan 1, sehingga variabel respon akan mengikuti distribusi Bernoulli untuk setiap observasi tunggal dengan fungsi probabilitas sebagai berikut,

$$f(y_i) = \pi(x_i)^{y_i}(1 - \pi(x_i))^{1-y_i} \quad ; y_i = 0, 1$$

Menurut Hosmer dan Lemeshow (2000) model umum regresi logistik dengan  $p$  faktor ( $p$  variabel prediktor) adalah

$$\pi(x_i) = \frac{e^{(\beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi})}}{1 + e^{(\beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi})}}$$

Untuk mempermudah pendugaan parameter regresi maka model regresi logistik dapat diuraikan dengan menggunakan transformasi logit dari  $\pi(x_i)$  sebagai berikut

$$\ln\left(\frac{\pi(x_i)}{1 - \pi(x_i)}\right) = \beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi}$$

Pendugaan parameter dalam model regresi logistik biner menggunakan metode kemungkinan maksimum (*Maximum Likelihood Estimation*), kemudian diselesaikan dengan metode iterasi Newton Raphson. Pengujian terhadap parameter-parameter model regresi logistik biner dilakukan baik secara simultan maupun secara parsial. Menurut Hosmer dan Lemeshow (2000), statistik uji yang digunakan untuk pengujian parameter model secara simultan adalah statistik uji  $G$  atau *likelihood ratio test*. Hipotesis dalam uji simultan adalah

$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$  (tidak ada pengaruh antara variabel prediktor dengan variabel respon)

$H_1 : \text{minimal ada satu } \beta_j \neq 0; j = 1, 2, \dots, p$   
(ada pengaruh paling sedikit satu variabel prediktor dengan variabel respon)

Statistik uji :

$$G = -2 \ln \frac{\binom{n_1}{n_1} \binom{n_0}{n_0}^{n_0}}{\sum_{i=1}^n \hat{\pi}_i^{y_i} (1 - \hat{\pi}_i)^{(1-y_i)}}$$

Kriteria penolakan  $H_0$  yaitu tolak  $H_0$  jika  $G > \chi^2(\alpha, v)$  dengan  $v$  derajat bebas adalah banyaknya parameter dalam model tanpa  $\beta_0$  atau nilai  $p - value < \alpha$ .

Setelah dilakukan pengujian parameter secara simultan, langkah selanjutnya adalah melakukan pengujian parameter secara parsial. Menurut Hosmer dan Lemeshow (2000), pengujian parameter secara parsial menggunakan uji Wald dengan hipotesis sebagai berikut:

$H_0 : \beta_j = 0$  (tidak ada pengaruh antara variabel prediktor dengan variabel respon) (2)

$H_1 : \beta_j \neq 0; j = 1, 2, \dots, p$  (ada pengaruh antara variabel prediktor dengan variabel respon)

Statistik uji :

$$W = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \quad (3)$$

Kriteria penolakan  $H_0$  yaitu tolak  $H_0$  jika  $W^2 > \chi^2(\alpha, v)$  dengan  $v$  merupakan derajat bebas,  $v = 1$  atau nilai  $p - value < \alpha$ .

Uji kesesuaian model digunakan untuk mengetahui apakah ada perbedaan yang signifikan antara hasil pengamatan dengan kemungkinan hasil prediksi model (Hosmer dan Lemeshow, 2000). Hipotesis yang digunakan adalah

$H_0 : \text{Model sesuai (tidak terdapat perbedaan yang signifikan antara hasil pengamatan dengan kemungkinan hasil prediksi model)}$

$H_1 : \text{Model tidak sesuai (terdapat perbedaan yang signifikan antara hasil pengamatan dengan kemungkinan hasil prediksi model)}$

Statistik Uji :

$$\hat{C} = \sum_{k=1}^g \frac{(o_k - n'_k \bar{\pi}_k)^2}{n'_k \bar{\pi}_k (1 - \bar{\pi}_k)}$$

Intepretasi koefisien untuk model regresi logistik biner dapat dilakukan dengan menggunakan nilai rasio oddsnya. Odds sendiri dapat diartikan sebagai rasio peluang kejadian sukses dengan kejadian tidak sukses dari peubah respon. Rasio odds didefinisikan sebagai :

$$\psi = \exp(\beta_i)$$

Menurut Hosmer dan Lemeshow (2000), salah satu ukuran kebaikan model dalam regresi logistik adalah jika memiliki peluang kesalahan klasifikasi yang minimal.

$$\text{Ketepatan Klasifikasi (\%)} = \left( \frac{n_{00} + n_{11}}{n_0 + n_1} \right) \times 100\%$$

$$\text{APER (\%)} = \left( \frac{n_{01} + n_{10}}{n_0 + n_1} \right) \times 100\%$$

### Analisis CART (*Classification and Regression Trees*)

CART menggunakan pendekatan nonparametrik, sehingga tidak ada asumsi harus mengikuti suatu distribusi tertentu, baik untuk variabel respon ataupun variabel prediktor. Selain itu, analisis ini tidak terpengaruh oleh adanya pencilan (*outlier*), kolinieritas, ataupun heterokedastisitas (Yohannes dan Webb, 1999). Menurut Breiman dkk. (1993), CART menghasilkan suatu pohon klasifikasi (*classification trees*) jika variabel responnya katagorik dan menghasilkan pohon regresi (*regression trees*) jika variabel responnya kontinu. Tujuan utama CART adalah untuk mendapatkan suatu kelompok data yang akurat sebagai penciri dari suatu pengklasifikasian.

#### Pembentukan Pohon Klasifikasi

Menurut Breiman dkk. (1993) pembentukan pohon klasifikasi terdiri atas 3 tahapan yaitu sebagai berikut :

##### A. Pemilahan Pemilah

Pemilihan pemilah menggunakan data *learning sample* ( $\mathcal{L}$ ) yang masih bersifat heterogen untuk pembentukan pohon klasifikasi. Pemilahan dilakukan untuk mengurangi heterogenan pada simpul utama dan memaksimumkan ukuran kehomogenan pada setiap simpul dalam. Untuk mengukur tingkat heterogenan suatu kelas dari suatu simpul pada pohon klasifikasi digunakan *Impurity measure*  $i(t)$ . Fungsi

*impurity measure* yang digunakan dalam penelitian ini adalah kaidah *Twoing*. Menurut Zambon dkk. (2006), fungsi kaidah *Twoing* dapat ditulis sebagai berikut:

$$i(t) = \frac{P_L P_R}{4} (\sum_i |(P(i|t_L) - P(i|t_R))|)^2$$

Misalkan terdapat pemilahan  $s$  yang memilah simpul  $t$  menjadi simpul kiri  $t_L$  (dengan proporsi  $P_L$ ) dan simpul kanan  $t_R$  (dengan proporsi  $P_R$ ) dimana  $P_R = 1 - P_L$  dan  $t_R \cup t_L = 1$ , maka kebaikan dari pemilah (*goodness of split*) didefinisikan sebagai penurunan heterogenan (*improvement*) (Breiman dkk., 1993) :

$$\Delta i(s, t) = i(t) - p_L \cdot i(t_L) - p_R \cdot i(t_R)$$

dimana :

$\Delta i(s, t)$ : penurunan tingkat heterogenan (*improvement*)

Pengembangan pohon dilakukan dengan cara pada simpul utama  $t_1$ , carilah  $s^*$  yang memberikan nilai penurunan heterogenan tertinggi, yaitu :

$$\Delta i(s^*, t_1) = \max_{s \in S} \Delta i(s, t_1)$$

Maka simpul  $t_1$  dipilah menjadi  $t_2$  dan  $t_3$  menggunakan  $s^*$ , dengan cara yang sama dilakukan juga pemilah terbaik pada  $t_2$  dan  $t_3$  secara terpisah dan seterusnya.

##### B. Penentuan Simpul Terminal

Suatu simpul  $t$  akan menjadi simpul terminal atau tidak akan dipilah lagi bila tidak terdapat penurunan heterogenan yang berarti, hanya terdapat satu pengamatan pada tiap simpul anak atau adanya batasan minimum  $n$  dan adanya batasan jumlah level atau tingkatan kedalaman pohon maksimal (Lewis, 2000).

##### C. Penandaan Label Kelas

Penandaan label kelas pada simpul terminal dilakukan berdasarkan aturan jumlah terbanyak, yaitu jika:

$$p(j_0|t) = \max_j p(j|t) = \frac{\max_j N_j(t)}{N(t)}$$

Maka label kelas untuk simpul terminal  $t$  adalah  $j_0$  yang memberikan nilai dugaan kesalahan pengklasifikasian pada simpul  $t$  terkecil sebesar

$$r(t) = 1 - \max_j p(j|t)$$

*Resubstitution estimate* (proporsi kesalahan) untuk tingkat kesalahan pengklasifikasian keseluruhan dari pohon klasifikasi, yaitu:

$$R(T) = \sum_{t \in T} R(t) = \sum_{t \in T} r(t) p(t)$$

### Penentuan Pohon Klasifikasi Optimal

Menurut Breiman dkk. (1993), salah satu cara mendapatkan pohon optimal yaitu dengan pemangkasan (*pruning*). Bagian pohon yang kurang penting dipangkas untuk mendapatkan pohon klasifikasi yang optimal. Ukuran pemangkasan yang digunakan untuk memperoleh ukuran pohon yang layak adalah *cost complexity measure*.

**Definisi 1.** ( Breiman dkk., 1993) Diketahui subtree  $T < T_{max}$  didefinisikan *complexity* (kompleksitas) dari subtree ini adalah  $|\tilde{T}|$ , yaitu banyaknya simpul terminal yang dimiliki  $T$ ,  $\alpha \geq 0$  adalah kompleksitas parameter dan *cost complexity measure* (ukuran biaya kompleksitas)  $R_\alpha(T)$  maka

$$R_\alpha(T) = R(T) + \alpha |\tilde{T}|$$

Proses pemangkasan pohon klasifikasi dimulai dengan mengambil  $t_R$  yang merupakan simpul anak kanan dan  $t_L$  yang merupakan simpul anak kiri dari  $T_{max}$  yang dihasilkan dari simpul induk  $t$ . Jika diperoleh dua simpul anak dan simpul induk yang memenuhi persamaan (2.31) maka simpul  $t_R$  dan  $t_L$  dipangkas. Hasilnya adalah pohon  $T_1$  yang memenuhi kriteria  $R(T_1) = R(T_{max})$ . Proses tersebut diulang sampai tidak ada lagi pemangkasan yang mungkin terjadi.

$$R(t) = R(t_R) + R(t_L)$$

Hasil proses pemangkasan berupa sederet pohon klasifikasi  $T_k$ . Pemilihan Pohon klasifikasi optimal berdasarkan pada *subtree* yang mempunyai nilai *cost complexity measure* yang terkecil dan nilai penduga pengganti yang cukup kecil. Nilai penduga pengganti diperoleh dengan menggunakan validasi silang (*cross-validation sample*) sebagai berikut:

$$R^{CV}(T_k) = \frac{1}{N} \sum_i C(i|j) N_{ij}$$

### Interpretasi Pohon Klasifikasi Optimal

Interpretasi dari pohon klasifikasi yang telah terbentuk disajikan dalam suatu aturan hubungan (*association rules*) yang disebut dengan *classification rules*. *Classification Rules* adalah aturan pengklasifikasian yang berbentuk *if...then...* (**jika... maka...**) (Breiman dkk, 1993).

### Ketepatan Prediksi Pohon Klasifikasi Optimal

Pohon klasifikasi optimal yang dihasilkan oleh data *learning* diuji tingkat keakuratannya dengan memasukkan data *testing* ke dalam pohon klasifikasi optimal, sehingga dihasilkan angka ketepatan klasifikasi. Uji keakuratan pohon klasifikasi optimal dilakukan dengan menghitung proporsi *misclassification*.

$$R(d) = \frac{1}{N} \sum_{n=1}^N X(d(x_n) \neq j_n)$$

Semakin besar  $R(d)$  maka tidak akurat pohon klasifikasi yang terbentuk dalam melakukan prediksinya. Ketepatan klasifikasi pohon klasifikasi optimal dapat diperoleh sebagai berikut

$$\text{Ketepatan Klasifikasi (\%)} = 1 - R(d)$$

### III. METODE PENELITIAN

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari data rekam medis pasien penyakit jantung yang menjalani rawat jalan di RSUD Dr. M. Yunus Bengkulu Tahun 2014. Sampel dalam penelitian ini sebanyak 846 orang. variabel respon ( $Y$ ) dalam penelitian ini adalah penyakit jantung yang terdiri dari dua kategori yaitu PJK dan NONPJK. Sedangkan variabel prediktor yang digunakan adalah jenis kelamin  $X_1$ , umur  $X_2$ , dislipidemia  $X_3$ , hipertensi  $X_4$ , diabetes melitus  $X_5$ , dan obesitas  $X_6$ .

Tahapan –tahapan yang akan dilakukan dalam penelitian ini:

1. Mendiskripsikan variabel - variabel penelitian yang digunakan
2. Analisis Regresi Logistik Biner. Tahapannya adalah sebagai berikut:
  - a. Melakukan pengujian multikolinieritas

- b. Membentuk model logit regresi logistik biner
  - c. Menduga parameter dengan metode maksimum *likelihood* dan iterasi Newton-Raphson.
  - d. Melakukan pengujian parameter secara serentak dengan uji G (*likelihood ratio test*)
  - e. Melakukan pengujian parameter secara parsial dengan uji *Wald*
  - f. Memilih model terbaik dengan uji Hosmer and Lemeshow
  - g. Interpretasi model dengan menggunakan *Odds ratio*
  - h. Menghitung ketepatan klasifikasi model
3. Analisis CART. Tahapannya adalah sebagai berikut:
- a. Membagi data menjadi 2 yaitu data *learning* dan data *testing*
  - b. Pembentukan pohon klasifikasi menggunakan data *learning*.
    - i. Mengukur tingkat keheterogenan untuk setiap variabel prediktor dengan menggunakan kaidah Twoing.
    - ii. Mengukur penurunan tingkat keheterogenan setiap variabel prediktor dengan menggunakan *Goodness of Split*.
    - iii. Memilih variabel dengan nilai *improvement* atau penurunan tingkat keheterogenan tertinggi sebagai kriteria dalam memilih pemilah terbaik yang digunakan untuk memecah sebuah simpul.
    - iv. Penentuan simpul terminal
    - v. Penandaan label kelas pada simpul terminal
  - c. Pembentukan pohon klasifikasi optimal dengan pemangkasan menggunakan *Cost Complexity Measure* dan *Cross Validation*.
  - d. Menguji keakuratan model pohon klasifikasi optimal yang dihasilkan oleh data *learning* dengan memasukkan data *testing* ke dalam model tersebut sehingga dihasilkan angka ketepatan klasifikasi.
  - e. Interpretasi model pohon klasifikasi optimal.

4. Membandingkan hasil dari analisis regresi logistik biner dan analisis CART berdasarkan tingkat ketepatan klasifikasi yang dihasilkan kedua analisis.

#### IV. HASIL DAN PEMBAHASAN

##### Deskripsi Data Pasien Penyakit Jantung yang Dirawat di RSUD Dr. M. Yunus Bengkulu

Pasien penyakit jantung yang dirawat di RSUD Dr. M. Yunus Bengkulu pada tahun 2014 ada lah sebanyak 846 orang. Jumlah pasien yang termasuk kategori PJK sebanyak 577 orang (68.2%) dan kategori NONPJK sebanyak 269 orang (31.8%). Pasien penyakit jantung mayoritas berjenis kelamin laki-laki, umur lebih besar dari 45 tahun, menderita dislipidemia, hipertensi, diabetes melitus dan obesitas.

##### Hasil Analisis Regresi Logistik Biner

Pendugaan model regresi logistik dengan enam variabel prediktor menghasilkan nilai *likelihood ratio* ( $G^2$ ) = 402.643 dengan nilai  $p$ -value = 0.000. Pengambilan keputusan dilakukan dengan membandingkan nilai  $G^2$  dengan kai-kuadrat dari tabel  $\chi^2_{(6;0.05)} = 12.592$ . Karena  $G^2 = 402.643 > \chi^2_{(6;0.05)} = 12.592$  dan  $p$ -value = 0.000 <  $\alpha = 0.05$  maka  $H_0$  ditolak, artinya minimal ada satu variabel prediktor yang berpengaruh secara simultan terhadap variabel respon. Setelah dilakukan pengujian signifikansi parameter secara simultan akan dilakukan pengujian secara parsial dengan menggunakan uji *Wald*. Pengujian signifikansi parameter secara parsial menghasilkan empat variabel prediktor yang berpengaruh nyata terhadap variabel respon yaitu dislipidemia, hipertensi, diabetes melitus dan obesitas. Model terbaik diperoleh dengan hanya memasukkan variabel yang signifikan ke dalam langkah analisisnya.

Berdasarkan Tabel 1(Lampiran) kolom *Sig.* diketahui bahwa variabel dislipidemia, diabetes melitus dan obesitas mempunyai nilai  $p$ -value = 0.000 <  $\alpha = 0.05$  dan variabel hipertensi mempunyai nilai  $p$ -value = 0.012 <  $\alpha = 0.05$ , artinya variabel dislipidemia, hipertensi, diabetes melitus dan

obesitas merupakan variabel yang signifikan berpengaruh terhadap penyakit jantung.

Dengan demikian model regresi logistik layak digunakan untuk memprediksi variabel respon dan model akhir logit dapat dinyatakan sebagai berikut:

$$g(x_i) = -3.163 + 2.482x_{3i} - 0.523x_{4i} + 2.435x_{5i} + 2.647x_{6i}$$

Sehingga diperoleh model terbaik regresi logistik biner sebagai berikut:

$$\pi(x_i) = \frac{\exp[g(x_i)]}{1 + \exp[g(x_i)]} = \frac{\exp[-3.163 + 2.482x_{3i} - 0.523x_{4i} + 2.435x_{5i} + 2.647x_{6i}]}{1 + \exp[-3.163 + 2.482x_{3i} - 0.523x_{4i} + 2.435x_{5i} + 2.647x_{6i}]}$$

Dari Tabel 1 dapat dilihat nilai rasio odds untuk variabel dislipidemia, hipertensi, diabetes melitus dan obesitas pada kolom Exp(B). Nilai rasio odds untuk variabel dislipidemia adalah 11.962, artinya penderita dislipidemia berisiko menderita PJK 11.962 kali dibandingkan yang bukan penderita dislipidemia. Variabel hipertensi mempunyai nilai rasio odds sebesar 0.592, artinya penderita hipertensi berisiko menderita PJK 0.592 kali dibandingkan yang bukan penderita hipertensi. Nilai rasio odds untuk variabel diabetes melitus adalah 11.421, artinya penderita diabetes melitus dibandingkan yang tidak diabetes melitus berisiko menderita PJK 11.421 kali. Variabel obesitas mempunyai nilai rasio odds sebesar 14.112, artinya penderita obesitas berisiko menderita PJK 14.112 kali dibandingkan yang bukan penderita obesitas.

Berdasarkan Tabel 2 (Lampiran) diketahui bahwa ketepatan klasifikasi yang dihasilkan metode regresi logistik biner adalah 84.2%.

## Hasil CART

Pohon klasifikasi maksimal yang dihasilkan mempunyai 28 simpul anak dan 15 simpul terminal dengan 7 kedalaman, dimana seluruh variabel prediktor berpengaruh pada pohon klasifikasi maksimal. Pohon klasifikasi maksimal yang terbentuk mengandung ukuran biaya kompleksitas (*test set relative cost*) sebesar  $0.12000 \pm 0.06499$  atau antara 0.05501 sampai 0.18499 dengan penduga pengganti (*resubstitution relative cost*) sebesar 0.32383

dan kompleksitas parameter 0.000000. Pohon klasifikasi maksimal menghasilkan nilai *relative cost* menggambarkan struktur data yang sangat kompleks, sehingga perlu dilakukan pemangkasan pohon agar diperoleh pohon optimal yang memiliki nilai *relative cost* yang lebih kecil. Pohon optimal yang terbentuk terdiri dari 6 simpul anak dan 4 simpul terminal dengan kedalaman 4 (Gambar 1 pada Lampiran). Pohon klasifikasi ukuran biaya kompleksitas (*test set relative cost*) sebesar  $0.10667 \pm 0.07539$  atau antara 0.03128 sampai 0.18206 dengan penduga pengganti (*resubstitution relative cost*) sebesar 0.33492 dan kompleksitas parameter 0.002371.

Adapun interpretasi hasil untuk pohon klasifikasi optimal sebagai berikut:

1. Jika pasien menderita dislipidemia maka diprediksi akan menderita PJK.
2. Jika pasien tidak menderita dislipidemia dan tidak menderita obesitas maka diprediksi akan menderita NONPJK.
3. Jika pasien tidak menderita dislipidemia tetapi menderita obesitas dan menderita diabetes melitus maka diprediksi akan menderita PJK.
4. Jika pasien tidak menderita dislipidemia tetapi menderita obesitas dan tidak menderita diabetes melitus maka diprediksi akan menderita NONPJK.

Berdasarkan Tabel 3 (Lampiran) diketahui bahwa ketepatan klasifikasi yang dihasilkan metode CART adalah 95%.

## V. KESIMPULAN DAN SARAN

### Kesimpulan

Berdasarkan hasil dan pembahasan yang telah diuraikan pada bab sebelumnya, maka diperoleh kesimpulan sebagai berikut:

1. Variabel prediktor yang signifikan terhadap variabel respon dari hasil analisis regresi logistik biner yaitu dislipidemia, hipertensi, diabetes melitus dan obesitas. Model regresi logistik biner yang diperoleh adalah

$$\pi(x_i) = \frac{\exp[-3.163 + 2.482x_{3i} - 0.523x_{4i} + 2.435x_{5i} + 2.647x_{6i}]}{1 + \exp[-3.163 + 2.482x_{3i} - 0.523x_{4i} + 2.435x_{5i} + 2.647x_{6i}]}$$

dimana  $x_i = 0$  dan  $1 ; i = 0,1$

Ketepatan klasifikasi yang dihasilkan metode regresi logistik biner adalah 84.2%.

2. Pohon klasifikasi yang terbentuk menghasilkan pohon optimal yang terdiri dari 6 simpul anak dan 4 simpul terminal dengan kedalaman 4. Variabel prediktor yang masuk ke dalam pohon klasifikasi yaitu dislipidemia, diabetes melitus dan obesitas. Variabel yang menjadi pemilah utama adalah variabel dislipidemia. Metode CART memiliki nilai ketepatan klasifikasi sebesar 95%.
3. Perbandingan hasil dari kedua analisis menunjukkan bahwa pada kasus ini metode CART lebih baik dari metode regresi logistik biner, karena metode CART memiliki ketepatan klasifikasi sebesar 95% sedangkan metode regresi logistik biner menghasilkan ketepatan klasifikasi sebesar 84.2%

## Saran

Penelitian tentang PJK ini hanya menggunakan enam variabel prediktor, yaitu variabel jenis kelamin, umur, dislipidemia, hipertensi, diabetes melitus dan obesitas. Penelitian selanjutnya hendaknya menggunakan variabel prediktor yang lebih lengkap, misalnya dengan menambahkan variabel keturunan, stress dan merokok.

## DAFTAR PUSTAKA

- Breiman, L., J. H. Friedman, R. A. Olshen and C. J. Stone. 1993. *Cassification and Regression Tree*. Chapman And Hall. New York.
- Gujarati, D. 2010. *Ekonometri Dasar. Selemba Empat*. Jakarta.
- Hosmer, D. W. and S. Lemeshow. 2000. *Applied Logistic Regression Second Edition*. John Willey & Son, Inc : New York.
- Lewis, R. J. 2000. *An Introduction to Classification and Regression Tree (CART) Analysis*. Presented at the 2000 Annual Meeting of the Society for Academic Emergency Medicine in San Francisco, California.

- Margasari, A. 2014. Penerapan Metode CART (Classification and Regression Trees) dan Analisis Regresi Logistik Biner pada Klasifikasi Profil Mahasiswa FMIPA Universitas Brawijaya. *Jurnal Mahasiswa Statistik*, Vol. 2, No. 4 (2014). FMIPA Universitas Brawijaya

<http://statistik.studentjournal.ub.ac.id/index.php/statistik/article/view/151>.

Diakses tanggal 16 Februari 2014.

- Neter, J., W. Wasserman and M. H. Kutner. 1990. *Applied Linear Statistical Models*. Third Edition. Richard D. Irwin, Inc, Homewood, Illinois 60430.
- Sezgin, O., E. Buyukbecici, A. I. Cekic, F. Y. Ozkurt, P. Taylan and G. H. Weber. 2006. *CMARS and GAM & CPQ, Modern Optimization Methods Applied to International Credit Default Prediction*. The Middle East Technical University. Turkey.
- Soeharto, I. 2004. *Serangan Jantung dan Stroke Hubungannya dengan Lemak dan Kolesterol Edisi Kedua*. Penerbit PT. Gramedia Pustaka Utama. Jakarta.
- Yohannes, Y. and P. Webb. 1999. *Classification and Regression Trees, CART<sup>TM</sup> : A User Manual For Identifying Indicators of Vulnerability to Famine and Chronic Food Insecurity*. International Food Policy Research Institute. Washington, U. S. A.
- Zakiyah, W., H. Permadi dan S. Rahardjo. 2012. Implementasi Regresi Logistik Biner pada Penentuan Faktor-faktor yang Mempengaruhi Penyakit Jantung. *Jurnal Online*, Vol. 1, No.3 (2012). Universitas Negeri Malang. <http://jurnal-online.um.ac.id/data/artikel/artikel202375F42424FC943D2F28AE7B51DA32.pdf> . Diakses tanggal 14 Februari 2015.
- Zambon, M., R. Lawrence, A. Bunn and S. Powell. 2006. Effect of Alternative Splitting Rules on Image Processing Using Classification Tree Analysis. *Photogrammetric Engineering & Remote Sensing* Vol. 72, N0. 1 : 25-30.

## LAMPIRAN

Tabel 1. Output Model Regresi Logistik Biner dengan Variabel Prediktor yang Signifikan

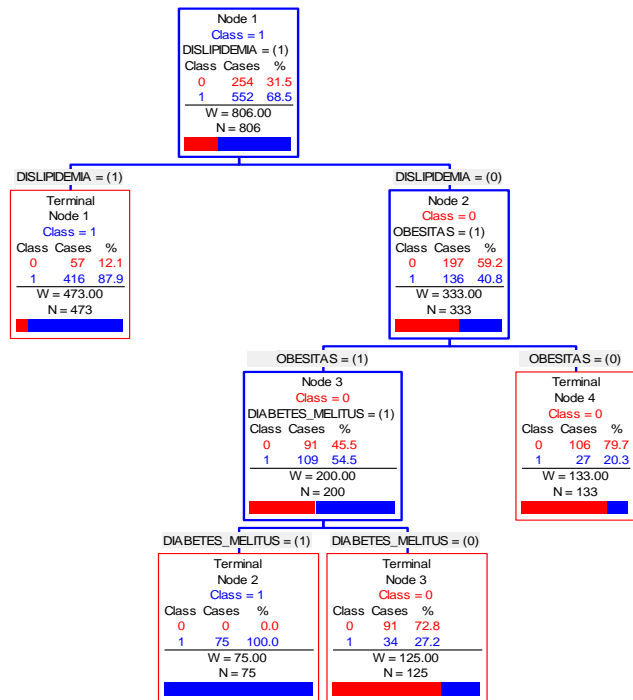
	B	S.E.	Wald	Df	Sig.	Exp(B)	95% C.I.for EXP(B)		
							Lower	Upper	
Step 1 <sup>a</sup>	Dislipidemia(1)	2.482	.207	144.347	1	.000	11.962	7.979	17.931
	Hipertensi(1)	-.523	.208	6.319	1	.012	.592	.394	.891
	Diabetes_Melitus(1)	2.435	.260	87.938	1	.000	11.421	6.865	19.000
	Obesitas(1)	2.647	.262	102.214	1	.000	14.112	8.448	23.576
	Constant	-3.163	.317	99.406	1	.000	.042		

a. Variable(s) entered on step 1: Dislipidemia, Hipertensi, Diabetes\_Melitus, Obesitas.

Tabel 2. Klasifikasi Analisis Regresi Logistik Biner

	Observed	Predicted		
		Penyakit_Jantung		Percentage Correct
		NONPJK	PJK	
Step 1	Penyakit_Jantung NONPJK	214	55	79.6
	PJK	79	498	86.3
	Overall Percentage			84.2

a. The cut value is .500



Gambar 1. Pohon Klasifikasi Optimal

Tabel 3. Ketepatan Klasifikasi Data *Testing* pada Pohon Optimal

Actual Class	Predicted		Total Class	Percent Correct
	NONPJK	PJK		
NONPJK	14	1	15	93.33%
PJK	1	24	25	96.00%
Overall Percent Correct				95%

## Prosedur *Multivariate Analysis of Variance* (MANOVA) Pada Regresi Multivariat

Harjuno Sosro<sup>\*</sup>, Sigit Nugroho<sup>\*\*</sup>, Etis Sunandi<sup>\*\*</sup>

Jurusan Matematika, FMIPA

Universitas Bengkulu

<sup>\*</sup>Mahasiswa Jurusan Matematika Fakultas MIPA Universitas Bengkulu

<sup>\*\*</sup>Staf Pengajar Jurusan Matematika Fakultas MIPA Universitas Bengkulu

### ABSTRAK

Penelitian ini bertujuan untuk agar dapat memahami prosedur *Multivariate Analysis of Variance* (MANOVA) pada Regresi Multivariat. Model linier yang digunakan pada penelitian ini adalah model regresi linier multivariat. Penelitian ini diharapkan dapat menambah wawasan tentang statistika multivariat, analisis regresi, serta *Multivariate Analysis of Variance* (MANOVA). Metode penelitian yang digunakan yaitu studi literatur. Penelitian ini menjelaskan bagaimana membentuk tabel MANOVA pada regresi multivariat. Dalam mengestimasi parameter, metode yang digunakan yaitu metode kuadrat terkecil. Teladan penerapan yang dijelaskan yaitu sampel urin manusia. Terdapat tiga belas variabel yang diamati yaitu dua variabel prediktor dan sebelas variabel respon. Akan dilihat pengaruh 1) Volume(ml); 2) (specific gravity- 1)  $\times 10^3$  terhadap 1) PH; 2) Koefisien modifikasi *creatinine*; 3) Pigmen *creatinine*; 4) *Phosphate*; 5) Kalsium; 6) *Phosphorus*; 7) *Creatinine*; 8) Chloride; 9) Boron; 10) Choline; 11) Copper. Dari hasil perhitungan dapat disimpulkan adanya pengaruh 1) Volume(ml); 2) (specific gravity- 1)  $\times 10^3$  terhadap 1) PH; 2) Koefisien modifikasi *creatinine*; 3) Pigmen *creatinine*; 4) *Phosphate*; 5) Kalsium; 6) *Phosphorus*; 7) *Creatinine*; 8) Chloride; 9) Boron; 10) Choline; 11) Copper

Kata kunci: Regresi Multivariat, Metode Kuadrat Terkecil, MANOVA

### PENDAHULUAN

Analisis regresi merupakan analisis statistik yang mempelajari hubungan antara variabel respon dengan variabel prediktor (Kurniawati, 2011). Tujuan dari analisis regresi adalah menentukan sebuah model, dengan model ini dapat diprediksi nilai dari variabel respon yang belum diketahui (Yan dan Su, 2009). Berdasarkan hubungan kelinieran antar parameter dalam persamaan regresi, analisis regresi dapat dibagi menjadi dua yaitu regresi linier dan regresi nonlinier. Berdasarkan jumlah variabel prediktor dan respon yang dilibatkan dalam analisis regresi, pada saat analisis regresi hanya terdiri dari satu variabel respon, regresi linier dapat dibagi menjadi regresi linier sederhana dan regresi linier berganda.

Regresi linier multivariat merupakan kelanjutan dari regresi linier berganda yang mana keduanya sama-sama menafsirkan hubungan linier antara variabel prediktor dan variabel respon tertentu. Pada regresi multivariat tidak hanya variabel prediktor yang berkorelasi satu sama lain, komponen variabel respon juga berkorelasi (Izenman, 2008).

Regresi linier sederhana dan regresi linier berganda penarikan kesimpulannya dilakukan dengan menggunakan *Analysis of Variance* (ANOVA) sedangkan regresi multivariat penarikan kesimpulannya dilakukan dengan menggunakan *Multivariate Analysis of Variance* (MANOVA). *Multivariate Analysis of Variance* (MANOVA)

merupakan pengembangan dari *Analysis of Variance* (ANOVA) untuk situasi dimana ada beberapa variabel respon (Tabachnick dan Fidell, 2007). Pada penelitian ini penulis tertarik untuk memahami dan mengkaji prosedur *Multivariate Analysis of Variance* (MANOVA) pada regresi multivariat.

## REGRESI LINIER BERGANDA

### 1. Model Regresi Linier Berganda

Terdapat dua variabel dalam analisis regresi, yaitu variabel respon dan variabel prediktor. Regresi digunakan untuk menduga nilai-nilai satu variabel respon dari nilai variabel prediktor yang sudah diketahui atau diasumsikan ada hubungan dengannya. Apabila suatu regresi linier hanya menggunakan satu variabel prediktor maka regresi linier ini disebut regresi linier sederhana, sedangkan yang menggunakan lebih dari satu variabel prediktor disebut regresi linier berganda (Nugroho, 2008).

Adapun bentuk dari persamaan regresi menurut Johnson (2007) dalam matriks dapat dituliskan sebagai berikut:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_{n-1} \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & \cdots & X_{1,k-1} & X_{1k} \\ 1 & X_{21} & \cdots & X_{2,k-1} & X_{2k} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & X_{n-1,1} & \cdots & X_{n-1,k-1} & X_{n-1,k} \\ 1 & X_{n1} & \cdots & & X_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{k-1} \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_{n-1} \\ \varepsilon_n \end{bmatrix}$$

Regresi linier berganda terdiri dari lebih dari satu variabel prediktor  $X_1, X_2, \dots, X_k$ . Berdasarkan bentuk dari persamaan regresi diatas, maka model dari regresi linier berganda menurut Rawlings (1998) dapat dituliskan sebagai berikut:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} + \varepsilon_i$$

dengan  $Y_i$  adalah variabel respon pada pengamatan ke-i,  $\beta_0$  adalah intersep,  $\beta_k$  adalah koefisien regresi pada variabel prediktor ke-k,  $X_{ik}$  adalah variabel prediktor ke-k pada pengamatan ke-i, dan  $\varepsilon_i$  adalah galat dari pengamatan ke-i.

### 2. Estimasi Parameter Regresi Linier Berganda

Pada regresi linier berganda estimasi parameter yang biasa digunakan adalah metode kuadrat terkecil yaitu meminumkan jumlah kuadrat galat. Adapun estimator kuadrat terkecil dari  $\beta$  adalah

$$\begin{aligned} (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\hat{\beta} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \\ \hat{\beta} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \end{aligned}$$

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & X_{11} & \cdots & X_{1k} \\ 1 & X_{21} & \cdots & X_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & \cdots & X_{nk} \end{bmatrix}, \hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{bmatrix}$$

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} n & \sum_{i=1}^n X_{i1} & \cdots & \sum_{i=1}^n X_{ik} \\ \sum_{i=1}^n X_{i1} & \sum_{i=1}^n X_{i1}^2 & \cdots & \sum_{i=1}^n X_{i1}X_{ik} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n X_{ik} & \sum_{i=1}^n X_{i1}X_{ik} & \cdots & \sum_{i=1}^n X_{ik}^2 \end{bmatrix}$$

$$\mathbf{X}'\mathbf{Y} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_{11} & X_{21} & \cdots & X_{n1} \\ \vdots & \vdots & \ddots & \vdots \\ X_{1k} & X_{2k} & \cdots & X_{nk} \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n Y_i \\ \sum_{i=1}^n X_{i1}Y_i \\ \vdots \\ \sum_{i=1}^n X_{ik}Y_i \end{bmatrix}$$

### 3. Analysis of Variance (ANOVA) Pada Regresi Linier Berganda

Model regresi yang baik ditandai oleh tingginya koefisien determinasi yang mana nilai koefisien determinasi ini didapatkan dari ANOVA. Adapun ANOVA dalam regresi linier berganda meliputi jumlah kuadrat (JK), derajat bebas (db), dan kuadrat tengah (KT).

**Tabel 1.** Analysis of Variance (ANOVA) Pada Regresi Linier Berganda

Sumber Keragaman	JK	DB	KT	$F^*$
Model Regresi	$\boldsymbol{\beta}'\mathbf{X}'\mathbf{Y} - \left(\frac{1}{n}\right)\mathbf{Y}'\mathbf{J}\mathbf{Y}$	$k$	$\frac{JK[Model]}{k}$	$\frac{KT[Model]}{KT[Galat]}$
Galat	$\mathbf{Y}'\mathbf{Y} - \boldsymbol{\beta}'\mathbf{X}'\mathbf{Y}$	$n - (k + 1)$	$\frac{JK[Galat]}{n - (k + 1)}$	
Total	$\mathbf{Y}'\mathbf{Y} - \left(\frac{1}{n}\right)\mathbf{Y}'\mathbf{J}\mathbf{Y}$	$n - 1$		

### 4. Pengujian Hipotesis

Hubungan antara variabel respon dengan variabel prediktor dalam analisis regresi linier berganda dapat diuji dengan menggunakan Uji F. Menurut Kutner, dkk (2004) hipotesis-hipotesis Pada uji F dapat dituliskan:

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

$$H_1: \text{tidak semua } \beta_j (j = 1, \dots, k) = 0$$

Statistik uji yang digunakan dalam uji F adalah:

$$F^* = \frac{KT[Model]}{KT[Galat]}$$

Kriteria penolakannya adalah:

- Jika  $F^* \leq F_{tabel}(1 - \alpha; k; n - (k + 1))$ , maka kesimpulannya  $H_0$  diterima

- Jika  $F^* > F_{tabel}(1 - \alpha; k; n - (k + 1))$ , maka kesimpulannya  $H_0$  ditolak

Hipotesis nol menyatakan semua koefisien regresi sama dengan nol. Artinya apabila hipotesis ini terpenuhi tidak adanya pengaruh variabel prediktor terhadap variabel respon.

## **MULTIVARIATE ANALYSIS OF VARIANCE (MANOVA) PADA REGRESI MULTIVARIAT**

### **1. Model Regresi Multivariat**

Model regresi multivariat adalah suatu model regresi dengan lebih dari satu variabel respon yang saling berkorelasi dan satu atau lebih variabel prediktor (Aminuddin dkk, 2013). Model ini merupakan pengembangan dari model regresi linier berganda. Menurut Johnson (2007) model regresi multivariat dalam bentuk matriks dapat dituliskan sebagai berikut:

$$\mathbf{Y}_{(n \times l)} = \mathbf{X}_{(n \times (k+1))} \boldsymbol{\beta}_{((k+1) \times l)} + \boldsymbol{\varepsilon}_{(n \times l)}$$

dengan  $E(\boldsymbol{\varepsilon}_{(i)}) = 0$  dan  $\text{Cov}(\boldsymbol{\varepsilon}_{(i)}, \boldsymbol{\varepsilon}_{(j)}) = \sigma_{ij} \mathbf{I}$

dimana:

$$\mathbf{Y}_{(n \times l)} = \begin{bmatrix} Y_{11} & Y_{12} & \cdots & Y_{1l} \\ Y_{21} & Y_{22} & \cdots & Y_{2l} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{n1} & Y_{n2} & \cdots & Y_{nl} \end{bmatrix} = [\mathbf{Y}_{(1)} \quad \mathbf{Y}_{(2)} \quad \cdots \quad \mathbf{Y}_{(l)}]$$

$$\mathbf{X}_{(n \times (k+1))} = \begin{bmatrix} 1 & X_{11} & \cdots & X_{1,k-1} & X_{1k} \\ 1 & X_{21} & \cdots & X_{2,k-1} & X_{2k} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & X_{n-1,1} & \cdots & X_{n-1,k-1} & X_{n-1,k} \\ 1 & X_{n1} & \cdots & X_{n,k-1} & X_{nk} \end{bmatrix}$$

$$\mathbf{X}_{(n \times (k+1))} = \begin{bmatrix} 1 & X_{11} & \cdots & X_{1,k-1} & X_{1k} \\ 1 & X_{21} & \cdots & X_{2,k-1} & X_{2k} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & X_{n-1,1} & \cdots & X_{n-1,k-1} & X_{n-1,k} \\ 1 & X_{n1} & \cdots & X_{n,k-1} & X_{nk} \end{bmatrix}$$

$$\boldsymbol{\varepsilon}_{(n \times l)} = \begin{bmatrix} \varepsilon_{11} & \varepsilon_{12} & \cdots & \varepsilon_{1l} \\ \varepsilon_{21} & \varepsilon_{22} & \cdots & \varepsilon_{2l} \\ \vdots & \vdots & \ddots & \vdots \\ \varepsilon_{n1} & \varepsilon_{n2} & \cdots & \varepsilon_{nl} \end{bmatrix} = [\boldsymbol{\varepsilon}_{(1)} \quad \boldsymbol{\varepsilon}_{(2)} \quad \cdots \quad \boldsymbol{\varepsilon}_{(l)}]$$

Pada model di atas,  $k$  merupakan banyaknya variabel prediktor yang dilibatkan dalam model,  $l$  merupakan banyaknya variabel respon yang dilibatkan dalam model, dan  $n$  merupakan banyaknya pengamatan. Matriks  $\mathbf{Y}_{(n \times l)}$  merupakan matriks dari variabel respon dimana variabel respon yang dilibatkan dalam model adalah sebanyak  $l$  dan diamati sebanyak  $n$ . Matriks  $\mathbf{X}_{(n \times (k+1))}$  merupakan matriks dari variabel prediktor dimana variabel prediktor yang dilibatkan di dalam model adalah sebanyak  $k$  dan diamati sebanyak  $n$ . Parameter  $\boldsymbol{\beta}_{((k+1) \times l)}$  merupakan matriks dari parameter-parameter regresi dimana matriks ini jumlah barisnya sebanyak  $(k + 1)$  dan jumlah kolomnya

sebanyak  $l$ .  $\boldsymbol{\epsilon}_{(n \times l)}$  merupakan matriks dari galat, dimana matriks dari galat ini jumlah barisnya sebanyak  $n$  dan jumlah kolomnya sebanyak  $l$ .

## 2. Asumsi Model Regresi Multivariat

Untuk menyelesaikan masalah dengan menggunakan model regresi multivariat harus diketahui terlebih dahulu apa saja hal-hal yang harus dipenuhi supaya model ini dapat digunakan. Hal-hal inilah yang disebut dengan asumsi, dengan kata lain apabila salah satu dari asumsi tidak terpenuhi maka model regresi multivariat tidak dapat digunakan untuk menyelesaikan masalah yang sedang diteliti. Menurut Aminuddin (2013) asumsi-asumsi model regresi multivariat adalah sebagai berikut:

1. Galatnya bersifat independen
2. Galatnya berdistribusi normal multivariat
3. Galatnya memiliki matriks varian kovarian yang homogen

## 3. Estimasi Parameter Regresi Multivariat

Sama halnya pada regresi linier berganda, pada regresi multivariat estimator kuadrat terkecil  $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ , namun pada regresi multivariat karena matriks  $\hat{\boldsymbol{\beta}}$  dan matriks  $\mathbf{Y}$  terdiri dari banyak kolom maka menurut Rencher (2002) estimator kuadrat terkecil  $\hat{\boldsymbol{\beta}}$  pada regresi multivariat dapat dituliskan sebagai berikut:

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{y}_{(1)} : \mathbf{y}_{(2)} : \dots : \mathbf{y}_{(l)}) \\ &= [(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}_{(1)} : (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}_{(2)} : \dots : (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}_{(l)}] \\ &= [\hat{\boldsymbol{\beta}}_{(1)} : \hat{\boldsymbol{\beta}}_{(2)} : \dots : \hat{\boldsymbol{\beta}}_{(l)}]\end{aligned}$$

dimana:

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_{01} & \hat{\beta}_{02} & \dots & \hat{\beta}_{0l} \\ \hat{\beta}_{11} & \hat{\beta}_{12} & \dots & \hat{\beta}_{1l} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\beta}_{k1} & \hat{\beta}_{k2} & \dots & \hat{\beta}_{kl} \end{bmatrix}$$

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} n & \sum_{i=1}^n X_{i1} & \dots & \sum_{i=1}^n X_{ik} \\ \sum_{i=1}^n X_{i1} & \sum_{i=1}^n X_{i1}^2 & \dots & \sum_{i=1}^n X_{i1}X_{ik} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n X_{ik} & \sum_{i=1}^n X_{i1}X_{ik} & \dots & \sum_{i=1}^n X_{ik}^2 \end{bmatrix}$$

$$\mathbf{X}'\mathbf{Y} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ X_{11} & X_{21} & \dots & X_{n1} \\ \vdots & \vdots & \ddots & \vdots \\ X_{1k} & X_{2k} & \dots & X_{nk} \end{bmatrix} \begin{bmatrix} Y_{11} & Y_{12} & \dots & Y_{1l} \\ Y_{21} & Y_{22} & \dots & Y_{2l} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{n1} & Y_{n2} & \dots & Y_{nl} \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{i=1}^n Y_{i1} & \sum_{i=1}^n Y_{i2} & \cdots & \sum_{i=1}^n Y_{il} \\ \sum_{i=1}^n X_{i1}Y_{i1} & \sum_{i=1}^n X_{i1}Y_{i2} & \cdots & \sum_{i=1}^n X_{i1}Y_{il} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n X_{ik}Y_{ik} & \sum_{i=1}^n X_{ik}Y_{i2} & \cdots & \sum_{i=1}^n X_{ik}Y_{il} \end{bmatrix}$$

#### 4. Multivariate Analysis of Variance (MANOVA)

Pada dasarnya analisis ragam peubah ganda (*Multivariate Analysis of Variance* atau *MANOVA*) merupakan pengembangan lebih lanjut dari *Analysis of Variance (ANOVA)*, keduanya memiliki tujuan yang sama yaitu untuk menganalisis data dan menarik kesimpulan.

**Tabel 2.** *Multivariate Analysis of Variance (MANOVA)* pada Regresi Multivariat

Sumber Keragaman	Matriks JK	Db
Model Regresi	$\mathbf{H} = \hat{\boldsymbol{\beta}}' \mathbf{X}' \mathbf{X} \hat{\boldsymbol{\beta}}$	$k$
Galat	$\mathbf{E} = \mathbf{Y}' \mathbf{Y} - \hat{\boldsymbol{\beta}}' \mathbf{X}' \mathbf{X} \hat{\boldsymbol{\beta}}$	$n - k - 1$
Total	$\mathbf{E}_H = \mathbf{Y}' \mathbf{Y}$	$n - 1$

#### 5. Pengujian Hipotesis

Untuk menguji apakah secara keseluruhan parameter-parameter regresi tidak sama dengan nol, dapat dilakukan pengujian sebagai berikut:

Hipotesis:

$$H_0: \boldsymbol{\beta}_{11} = \boldsymbol{\beta}_{12} = \cdots = \boldsymbol{\beta}_{k1} = \cdots = \boldsymbol{\beta}_{kl} = \mathbf{0}$$

$$H_1: \text{Paling sedikit ada satu } \boldsymbol{\beta}_{kl} \neq \mathbf{0}$$

Statistik Uji yang digunakan adalah Wilk's Lambda:

$$\Lambda = \frac{|\mathbf{E}|}{|\mathbf{E} + \mathbf{H}|}$$

$H_0$  ditolak jika  $\Lambda_{hitung} \leq \Lambda_{\alpha, l, n-k-1}$  adalah nilai tabel kritis untuk Wilk's lambda.

Statistik wilk's lambda dapat ditransformasi menuju statistik F yang berdistribusi F Fisher. Sehingga dalam pengujian akan menggunakan uji F. adapun aturannya adalah sebagai berikut:

**Tabel 3.** Transformasi Wilk's Lambda ke distribusi F

Parameter		$F_{hitung}$	Db F
$l$	DB Model		
1	$\geq 1$	$\left( \frac{1 - \Lambda}{\Lambda} \right) \left( \frac{DB[G]}{DB[M]} \right)$	$Db[M], Db[G]$

2	$\geq 1$	$\left(\frac{1 - \sqrt{\Lambda}}{\sqrt{\Lambda}}\right) \left(\frac{DB[G] - 1}{DB[M]}\right)$	$2(Db[M]), 2((Db[G]) - 1)$
$\geq 1$	1	$\left(\frac{1 - \Lambda}{\Lambda}\right) \left(\frac{DB[G] + DB[M] - l}{l}\right)$	$l, Db[G] + Db[M] - l$
$\geq 1$	2	$\left(\frac{1 - \sqrt{\Lambda}}{\sqrt{\Lambda}}\right) \left(\frac{DB[G] + DB[M] - l - 1}{l}\right)$	$2l, 2(Db[G] + Db[M] - l - 1)$

dengan:

$l$  = Banyaknya variabel respon yang diamati

$Db[M]$  = Derajat bebas model regresi

$Db[G]$  = Derajat bebas galat

Setelah melakukan transformasi pengujian akan dilakukan dengan menggunakan uji F, dengan kriteria penolakan tolak  $H_0$  jika  $F_{Hitung} > F_{Tabel}$

## TELADAN PENERAPAN

Teladan penerapan *Multivariate Analysis of Variance* (MANOVA) pada regresi multivariat diambil dari teladan pada buku "*Multivariate Observations*" oleh Seber (1984). Data tersebut merupakan data sampel urin pada manusia. Dimana hal-hal yang diamati ada tiga belas variabel. Tiga belas variabel tersebut terdiri dari sebelas variabel respon dan dua variabel prediktor.

### 1. Mode Linier

Model regresi multivariat dapat dituliskan:

$$Y_{(45 \times 11)} = X_{(45 \times (2+1))} \beta_{((2+1) \times 11)} + \epsilon_{(45 \times 11)}$$

### 2. Estimasi Parameter

Menurut Rencher (2002) estimator kuadrat terkecil  $\hat{\beta}$  pada regresi multivariat dapat dituliskan sebagai berikut:

$$\hat{\beta} = (X'X)^{-1}X'Y$$

Jika ditulis dalam bentuk  $\hat{\beta}$  matriksnya adalah sebagai berikut:

$$\begin{bmatrix} 5,7876 & 4,8178 & 12,6788 & 1,8931 & -0,0545 & 2,2068 & 1,8464 & -2,1392 & 6,0826 & 0,5366 & 0,2616 \\ 0,0001 & -0,0066 & 0,0172 & -0,0046 & 0,0001 & -0,0033 & -0,0039 & 0,0043 & -0,0071 & -0,0069 & -0,0002 \\ -0,0119 & 0,0017 & -0,0508 & 0,0509 & 0,0081 & 0,0182 & 0,0583 & 0,2423 & -0,0436 & 0,3382 & -0,0018 \end{bmatrix}$$

### 3. *Multivariate Analysis of Variance* (MANOVA)

Pertama yang perlu ditemukan adalah nilai matriks jumlah kuadrat model regresi ( $H$ ). Adapun matriks jumlah kuadrat model regresi yaitu:

$$H = \tilde{\beta}' X' X \tilde{\beta}$$

Untuk mencari nilai matriks jumlah kuadrat galat yaitu:

$$E = Y'Y - \tilde{\beta}' X' X \tilde{\beta}$$

dengan

derajat bebas model regresi = 2  
 derajat bebas galat = 45 - 2 - 1 = 42  
 derajat bebas total = 45 - 1 = 44

**Tabel 4.** *Multivariate Analysis of Variance (MANOVA)* pada sample urin manusia

Sumber Keragaman	Matriks JK	Db
Model Regresi	$\begin{bmatrix} 1360,389 & 774,0005 & \dots & 39,52077 \\ 774,0005 & 464,6107 & \dots & 23,05335 \\ \vdots & \vdots & \ddots & \vdots \\ 39,52077 & 23,05535 & \dots & 1,164657 \end{bmatrix}$	2
Galat	$\begin{bmatrix} 3,6506 & 3,9334 & \dots & -0,132 \\ 3,9334 & 26,99 & \dots & -0,135 \\ \vdots & \vdots & \ddots & \vdots \\ -0,132 & -0,135 & \dots & 0,0982 \end{bmatrix}$	42
Total	$\begin{bmatrix} 1364,04 & 777,394 & \dots & 39,3890 \\ 777,934 & 491,601 & \dots & 22,9205 \\ \vdots & \vdots & \ddots & \vdots \\ 39,3890 & 22,9205 & \dots & 1,26290 \end{bmatrix}$	44

#### 4. Pengujian Hipotesis

Pengujian ini bertujuan untuk melihat apakah secara keseluruhan parameter-parameter regresi tidak sama dengan 0 artinya adanya pengaruh variabel prediktor terhadap variabel respon. Pengujiannya adalah sebagai berikut:

- Hipotesis:  
 $H_0: \beta_{11} = \beta_{12} = \dots = \beta_{21} = \dots = \beta_{2,11} = 0$   
 $H_1: \text{Paling sedikit ada satu } \beta_{kl} \neq 0$
- Besaran yang diperlukan:  
 $\alpha = 0,05$
- Statistik Uji:

$$\Lambda = \frac{|\mathbf{E}|}{|\mathbf{E} + \mathbf{H}|}$$

dimana:

$$\mathbf{E} = \begin{bmatrix} 3,6506 & 3,9334 & \dots & -0,132 \\ 3,9334 & 26,99 & \dots & -0,135 \\ \vdots & \vdots & \ddots & \vdots \\ -0,132 & -0,135 & \dots & 0,0982 \end{bmatrix}$$

$$|\mathbf{E}| = 544873768988,481$$

$$\mathbf{E} + \mathbf{H} = \begin{bmatrix} 1364,04 & 777,394 & \dots & 39,3890 \\ 777,934 & 491,601 & \dots & 22,9205 \\ \vdots & \vdots & \ddots & \vdots \\ 39,3890 & 22,9205 & \dots & 1,26290 \end{bmatrix}$$

$$|\mathbf{E} + \mathbf{H}| = 4825043144535070$$

sehingga

$$\Lambda = \frac{|\mathbf{E}|}{|\mathbf{E} + \mathbf{H}|} = \frac{544873768988,481}{4825043144535070} = 0,000112926$$

$$\Lambda_{Hitung} = 0,000112926$$

Setelah didapatkan nilai wilk's lambda, nilai tersebut bisa ditransformasi menuju sebaran F sehingga pengujiannya mengikuti uji F. Diketahui banyaknya variabel respon adalah sebelas serta derajat bebas modelnya 2, maka untuk mentransformasi menuju F mengikuti persamaan berikut:

$$F_{Hitung} = \left( \frac{1 - \sqrt{\Lambda}}{\sqrt{\Lambda}} \right) \left( \frac{DB[G] + DB[M] - l - 1}{l} \right)$$

$$F_{Hitung} = \left( \frac{1 - \sqrt{0,000112926}}{\sqrt{0,000112926}} \right) \left( \frac{42 + 2 - 11 - 1}{11} \right) = 270,8446$$

$$\begin{aligned} \text{dengan derajat bebas } F_{tabel} & 2l, 2(Db[G] + Db[M] - l - 1) \\ & = 2(11), 2(42 + 2 - 11 - 1) \\ & = 22; 64 \end{aligned}$$

artinya pada  $F_{tabel} Db[M] = 22, Db[G] = 64$ .

$$F_{0,05;22;64} = 1,71084$$

- Kriteria penolakan  
Tolak  $H_0$  jika  $F_{Hitung} > F_{Tabel}$   
Terima  $H_0$  jika  $F_{Hitung} \leq F_{Tabel}$
- Kesimpulan  
 $H_0$  ditolak, karena  $(F_{Hitung} = 270,8446) > (F_{Tabel} = 1,71084)$ , artinya Paling sedikit ada satu  $\beta_{kl} \neq 0$ . Hal ini menunjukkan secara bersama-sama Volume dan specific gravity mempengaruhi variabel-variabel respon yaitu PH, Koefisien modifikasi *creatinine*, Pigmen *creatinine*, *Phosphate*, Kalsium, *Phosphorus*, *Creatinine*, Chloride, Boron, Choline, Copper.

## KESIMPULAN

Model linier untuk regresi multivariat merupakan perluasan dari regresi linier berganda, yaitu:

$$\mathbf{Y}_{(n \times l)} = \mathbf{X}_{(n \times (k+1))} \boldsymbol{\beta}_{((k+1) \times l)} + \boldsymbol{\varepsilon}_{(n \times l)}$$

Metode kuadrat terkecil merupakan salah satu metode untuk mengestimasi parameter regresi. Parameter perlu diestimasi untuk memperkirakan nilai-nilai parameter tersebut. Adapun estimator kuadrat terkecil pada regresi multivariat adalah:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{y}_{(1)} \vdots \mathbf{y}_{(2)} \vdots \dots \vdots \mathbf{y}_{(l)})$$

Penggunaan tabel keragaman pada regresi multivariat bukan menggunakan ANOVA melainkan MANOVA. MANOVA merupakan pengembangan lebih lanjut dari ANOVA dalam kasus variabel respon yang diamati lebih dari satu. MANOVA pada regresi

multivariat memiliki matriks jumlah kuadrat model regresi  $\mathbf{H} = \hat{\boldsymbol{\beta}}' \mathbf{X}' \mathbf{X} \hat{\boldsymbol{\beta}}$ , matriks jumlah kuadrat galat  $\mathbf{E} = \mathbf{Y}' \mathbf{Y} - \hat{\boldsymbol{\beta}}' \mathbf{X}' \mathbf{X} \hat{\boldsymbol{\beta}}$ , serta matriks jumlah kuadrat total  $\mathbf{E}_{\mathbf{H}} = \mathbf{Y}' \mathbf{Y}$ .

## SARAN

Dalam penulisan skripsi ini, penulis hanya menjelaskan prosedur *Multivariate Analysis of Variance* (MANOVA) pada regresi multivariat dengan teladan penerapan yang diambil dalam buku Seber (1984) "*Multivariate Observations*". Untuk penelitian lebih lanjut, dapat dibahas mengenai prosedur pemilihan model terbaik pada regresi multivariat.

## DAFTAR PUSTAKA

- Aminuddin, Sudarno, dan Sugito. 2013. *Pemilihan Model Regresi Linier Multivariat Terbaik Dengan Kriteria Mean Square Error*. UNDIP.
- Gujarati, D.N. 2007. *Dasar-dasar Ekonometrika*. Erlangga.
- Izenman, A. J. 2008. *Modern Multivariate Statistical Techniques*. Springer. USA
- Johnson R.A. and D.W. Wichern. 2007. *Applied Multivariate Statistical Analysis 6th*. Pearson Education International. USA
- Kurniawan A. 2009. *Belajar mudah SPSS untuk Pemula*. MediaKom. Yogyakarta
- Kurniawati L.D. 2011. Kekekaran Regresi Linier Ganda Dengan Estimasi MM (Method of Moment) Dalam Mengatasi Pencilan. *Skripsi*. Universitas Negeri Yogyakarta. Yogyakarta.
- Kutner, M.H, J.N. Christopher, N. John, and L. William. 2004. *Applied Linear Statistical Models*. Fifth Edition. McGraw-Hill/Irwin. New York
- Mattjik, A.A dan I.M. Sumertajaya. 2011. *Sidik Peubah Ganda*. Intitut Pertanian Bogor. Bogor.
- Rawlings, J.O, S.G. Pantula dan D.A. Dickey. 1998. *Applied Regression Analysis: A Research Tool 2nd*. Springer. USA.
- Rencher, A.C. 1998. *Multivariate Statistical Inference and Applications*. Brigham Young University. Canada.
- Rencher, A.C. 2002. *Method of Multivariate Analisis*. Brigham Young University. Canada.
- Santoso, S. 2005. *Menggunakan SPSS untuk Statistik Parametrik*. Elex Media Komputindo. Jakarta.
- Seber. G. A. F. 1984. *Multivariate Observations*. John Wiley and Sons. New Zealand.
- Sigit Nugroho Ph.D. 2008. *Dasar-Dasar Metode Statistika*. Grasindo. Jakarta
- Sigit Nugroho Ph.D. 2008. *Dasar-Dasar Rancangan Percobaan*. Unib Press
- Tabachnick, B.G and L.S. Fidell. 2007. *Using Multivariate Statistics*. California State University. Northridge.
- Widiharih, T 2011, 'Analisis Ragam Multivariat Untuk Rancangan Acak Lengkap Dengan Pengamatan Berulang', *Jurnal Matematika dan Komputer*, vol. 4. No. 3, hh 139-150.
- Yan X. And X.G. Su. 2009. *Linear Regression Analysis Theory and Computing*. World Scientific. Singapore.

**REGRESI BINOMIAL NEGATIF TERBOBOTI GEOGRAFIS UNTUK MENGETAHUI FAKTOR-FAKTOR YANG MEMPENGARUHI JUMLAH KASUS MALARIA DI PROVINSI BENGKULU**  
**(GEOGRAPHICALLY WEIGHTED NEGATIVE BINOMIAL REGRESSION TO DETERMINE THE FACTORS THAT INFLUENCE THE NUMBER OF MALARIA IN BENGKULU PROVINCE)**

**Ummu Kholisoh Zuhry<sup>1</sup>, Sigit Nugroho<sup>2</sup>, Dyah Setyo Rini<sup>3</sup>**  
Jurusan Matematika, FMIPA, Universitas Bengkulu  
Jl. WR. Supratman Kandang Limun Bengkulu 38371A  
e-mail : [ummuzuhry13@gmail.com](mailto:ummuzuhry13@gmail.com)

**ABSTRACT**

This research will determine the factors that influence the number of malaria in Bengkulu Province in 2013. Geographically Weighted Negative Binomial Regression (GWNBR) can be used to overcome overdispersion in count data by considering the spatial effects, that is Adaptive Bisquare Kernel. The result of this research shows that Akaike Information Criterion (AIC) of GWNBR was smaller than Negative Binomial Regression, so GWNBR model is better. GWNBR model classifies regency/city into 4 groups based on the variables that significantly influence the number of malaria for each regency/city in Bengkulu Province.

**Keywords** : Overdispersion, Spatial Heterogeneity, Negative Binomial Regression, GWNBR, The number of malaria

**ABSTRAK**

Penelitian ini bertujuan untuk mengetahui faktor-faktor yang mempengaruhi jumlah kasus malaria di Provinsi Bengkulu tahun 2013 yang merupakan data cacah. Untuk mengatasi overdispersi pada data cacah dengan mempertimbangkan efek spasial digunakan Regresi Binomial Negatif Terboboti Geografis (RBNTG), dengan menggunakan salah satu fungsi pembobot yaitu *Adaptive Bisquare Kernel*. Hasil penelitian ini menunjukkan bahwa nilai AIC model RBNTG lebih kecil dari AIC model Regresi Binomial Negatif, mengindikasikan bahwa model RBNTG lebih tepat digunakan untuk data cacah yang mengalami overdispersi yang memperhatikan efek spasial karena dapat diketahui variabel yang berpengaruh signifikan untuk setiap lokasi pengamatan. Dari model Regresi Binomial Negatif Terboboti Geografis diperoleh 4 pengelompokan berdasarkan variabel yang berpengaruh signifikan sehingga dapat diketahui faktor apa saja yang mempengaruhi jumlah kasus malaria di setiap kabupaten/kota di Provinsi Bengkulu.

**Kata kunci**: Overdispersi, Heterogenitas spasial, Regresi Binomial Negatif, RBNTG, Jumlah kasus malaria

**A. PENDAHULUAN**

**1. Latar Belakang**

[1] menyatakan bahwa untuk melihat tingkat kesejahteraan rumah tangga suatu wilayah terdapat beberapa indikator yang dapat dijadikan sebagai ukuran, salah satunya adalah tingkat kesehatan.

Pembangunan kesehatan di Indonesia masih jauh dari yang diharapkan, hal ini terlihat dari banyaknya permasalahan kesehatan, salah satunya adalah angka kejadian malaria (*Annual Parasite*

*Incidence/API*). Diantara provinsi lain di Indonesia, Provinsi Bengkulu menempati urutan provinsi keenam dengan kasus terbesar malaria dengan angka kejadian malaria (*API/Annual Parasite Incidence*) sebesar 3,89%.

Jumlah kasus malaria merupakan data cacah (*count data*) dengan peluang kejadian kecil dan kejadiannya tergantung pada interval waktu tertentu atau berlangsung pada daerah tertentu. Pada data cacah kadangkala terjadi kondisi dimana nilai varian variabel respon lebih besar dari rata-ratanya yang disebut overdispersi [5].

Menurut [8], untuk memodelkan data cacah yang mengalami overdispersi, distribusi Binomial Negatif lebih cocok digunakan karena menurut [2] pada distribusi Binomial Negatif tidak mengharuskan nilai varian variabel responnya sama dengan rata-ratanya.

Keragaman spasial atau heterogenitas spasial yang terjadi disebabkan adanya efek spasial yang ditandai dengan perbedaan kondisi geografis, sosial budaya dan ekonomi antara wilayah satu dengan wilayah yang lain [7]. Menurut [8], masalah heterogenitas spasial tersebut mampu diatasi dengan salah satu metode hasil pengembangan model regresi linier lokal non parametrik yaitu metode GWR dengan menggunakan pendekatan titik yang akan menghasilkan pendugaan parameter lokal berdasarkan posisi atau jarak wilayah pengamatan satu dengan wilayah lainnya dengan diberikan pembobotan.

Oleh karena itu, untuk mengatasi masalah overdispersi pada data cacah dengan memperhatikan aspek spasial, [7] menyarankan menggunakan metode *Geographically Weighted Negative Binomial Regression* (Regresi Binomial Negatif Terboboti Geografis).

## **2. Rumusan Masalah**

1. Bagaimana memodelkan jumlah kasus malaria menggunakan Regresi Binomial Negatif Terboboti Geografis?
2. Faktor-faktor apa sajakah yang mempengaruhi jumlah kasus malaria di Provinsi Bengkulu dengan menggunakan Regresi Binomial Negatif Terboboti Geografis?

## **3. Tujuan Penelitian**

1. Memodelkan jumlah kasus malaria menggunakan Regresi Binomial Negatif Terboboti Geografis
2. Mengetahui faktor-faktor yang berpengaruh terhadap jumlah kasus malaria di Provinsi Bengkulu dengan menggunakan Regresi Binomial Negatif Terboboti Geografis

## B. METODE PENELITIAN

Data yang digunakan diperoleh dari Profil Kesehatan Provinsi Bengkulu 2013 dan Riset Kesehatan Dasar Provinsi Bengkulu 2013. Penelitian ini menggunakan data jumlah kasus malaria pada tahun 2013 di Provinsi Bengkulu yang terdiri dari 10 Kabupaten/Kota. Variabel yang digunakan untuk penelitian ini adalah:

- a. Jumlah kasus malaria di tiap kabupaten/kota di Provinsi Bengkulu tahun 2013 ( $Y + 1$ )
- b. Persentase rumah sehat pada tiap Kabupaten/Kota ( $X_1$ )
- c. Presentase rumah tangga berperilaku hidup bersih dan sehat pada tiap Kabupaten/Kota ( $X_2$ )
- d. Presentase pengobatan efektif dengan ACT pada tiap Kabupaten/Kota ( $X_3$ )
- e. Variabel spasial  $(u_i, v_i)$  yaitu titik koordinat lokasi dalam jarak (kilometer) masing-masing kabupaten/kota di Provinsi Bengkulu.

Tahapan analisis dalam penelitian ini adalah sebagai berikut:

1. Mengeksplorasi variabel dependen menggunakan peta tematik untuk memperoleh deskripsi data dengan bantuan program *ArcView GIS 3.2*
2. Melakukan uji asumsi multikolinearitas
3. Melakukan statistika deskriptif dan mendeteksi adanya kasus overdispersi
4. Melakukan analisis regresi Binomial Negatif dengan bantuan program *R 3.2.2*
5. Melakukan pengujian spasial yaitu uji heterogenitas spasial dengan bantuan program *R 3.2.2*
6. Melakukan analisis model RBNTG dengan bantuan program *R 3.2.2*
  - a. Menghitung jarak *Euclidean* antar lokasi pengamatan berdasarkan lokasi geografis
  - b. Menentukan *bandwidth* optimum berdasarkan kriteria CV yang minimum
  - c. Menghitung matriks pembobot menggunakan fungsi *kernel adaptive bisquare*
  - d. Menaksir parameter model RBNTG dengan metode MLE
  - e. Melakukan pengujian signifikansi parameter model RBNTG kemudian menghitung nilai AIC
7. Melakukan pengujian kesesuaian model dan melakukan perbandingan antara model Binomial Negatif dan model RBNTG

## C. HASIL DAN PEMBAHASAN

Provinsi Bengkulu merupakan salah satu provinsi yang berada di Pulau Sumatera dengan wilayah yang memanjang dari perbatasan Provinsi Sumatera Barat hingga ke perbatasan Provinsi Lampung dan jaraknya lebih kurang 567 kilometer. Luas wilayah Provinsi Bengkulu mencapai lebih

kurang 1.991.933 hektar atau 19.919,33 kilometer persegi. Secara astronomis, Provinsi Bengkulu terletak di antara 2°16' LU dan 3°31' LS dan antara 101°01' - 103°41' BT.

Tabel 1 menunjukkan bahwa variabel jumlah kasus malaria ( $Y_i + 1$ ) memiliki jangkauan yang sangat besar, yaitu 4133 sedangkan jangkauan yang paling kecil adalah variabel persentase pengobatan efektif dengan ACT ( $X_3$ ) sebesar 78,9. Jumlah kasus malaria terbesar yaitu sebanyak 4133 kasus berada di kabupaten Rejang Lebong.

**Tabel 1. Statistika deskripsi variabel dependen dan independen**

Variabel	Jangkauan	Minimum	Maksimum	Nilai tengah	Simpangan baku	Varian
$Y + 1$	4132	1	4133	763,90	1246,420	1553562,989
$X_1$	91	8	99	62,70	32,945	1085,344
$X_2$	85.4	0	85.4	52,220	23,1225	534,648
$X_3$	78.9	0	78.9	45,470	22,7447	517,322

### 1. Multikolinearitas

Nilai *Variance Inflation Factor* (VIF) merupakan nilai yang digunakan untuk mendeteksi multikolinearitas untuk mengetahui apakah terdapat hubungan antara beberapa atau semua variabel pada analisis regresi ini. Nilai VIF yang lebih dari 10 menunjukkan adanya multikolinearitas [4].

**Tabel 2. Nilai VIF variabel independen**

Variabel	$X_1$	$X_2$	$X_3$
Nilai VIF	1,471791	1,271824	1,249901

Tabel 2 menunjukkan bahwa nilai VIF dari  $X_1, X_2$  dan  $X_3$  lebih kecil dari 10, yang artinya tidak terdapat multikolinearitas antar variabel independen tersebut.

### 2. Overdispersi

Data jumlah kasus malaria memiliki varian yang besar yaitu 1553562.9 dibandingkan dengan nilai tengahnya yang hanya 763,90 (Tabel 4.2). [5] menyatakan bahwa nilai varian yang lebih besar daripada nilai tengah pada data cacah menunjukkan data mengalami overdispersi.

### 3. Model Regresi Binomial Negatif

Langkah awal pemodelan regresi Binomial Negatif yaitu penentuan nilai awal, diperoleh  $\hat{\theta} = 0,634$  selanjutnya didapat estimasi parameter model regresi Binomial Negatif (Tabel 4).

**a. Uji Simultan**

Hipotesis pengujian parameter secara simultan adalah:

$H_0: \beta_1 = \beta_2 = \beta_3 = 0$  (semua variabel secara bersama-sama memberikan pengaruh terhadap model)

$H_1: \beta_k \neq 0, k = 1,2,3$  (paling sedikit ada satu variabel yang memberikan pengaruh terhadap model)

Daerah penolakannya adalah  $D(\hat{\beta}) > \chi^2_{(\alpha;p)}$ .

**Tabel 3. Nilai Devians model regresi Binomial Negatif**

Devians	$\chi^2_{(0.05;3)}$
11,985	7.814

Nilai Devians lebih besar dari  $\chi^2_{(0.05;3)}$  yaitu  $11,985 > 7,814$ . Jadi  $H_0$  ditolak, artinya pada taraf nyata pengujian 5%, paling sedikit ada satu  $\beta_k \neq 0, k = 1,2,3$  yang berpengaruh signifikan terhadap variabel respon.

**b. Uji Parsial**

Hipotesis pengujian parameter secara parsial sebagai berikut:

$H_0: \beta_k = 0$  (variabel  $k$  tidak memberikan pengaruh yang signifikan terhadap variabel respon)

$H_1: \beta_k \neq 0$  dengan  $k = 1,2,3$  (variabel  $k$  memberikan pengaruh yang signifikan terhadap variabel respon)

Daerah penolakannya adalah  $|Z_{hit}| > Z_{\alpha/2}$ .

**Tabel 4. Estimasi parameter model regresi Binomial Negatif**

Parameter	Galat baku	Estimasi	$Z_{hitung}$	$Z_{\alpha/2}$	Keputusan
$\hat{\beta}_0$	1,37545	7,6226	5,542	1.959	$H_0$ ditolak
$\hat{\alpha}_1$	0,01560	0,0626	4,012	1.959	$H_0$ ditolak
$\hat{\beta}_2$	0,02069	-0,0748	-3,617	1.959	$H_0$ ditolak
$\hat{\beta}_3$	0,02064	-0,0383	-1,855	1.959	$H_0$ diterima

Berdasarkan Tabel 4 untuk masing-masing  $\hat{\beta}_1$  dan  $\hat{\beta}_2$  keputusannya adalah  $H_0$  ditolak karena nilai Zhitung lebih besar dari nilai  $Z_{\alpha/2}$ , sedangkan nilai Zhitung  $\hat{\beta}_3$  lebih kecil dari nilai  $Z_{\alpha/2}$  maka  $H_0$  diterima dengan taraf nyata pengujian sebesar 5%. Jadi, variabel independen yang signifikan mempengaruhi model adalah parameter  $\hat{\beta}_1$  dan  $\hat{\beta}_2$ . Model Regresi Binomial Negatif yang diperoleh adalah:

$$\mu = \exp(7,62255 + 0,0626X_1 - 0,07483X_2)$$

#### 4. Pengujian Heterogenitas Spasial

Untuk mengetahui adanya keragaman spasial pada data dapat dilakukan pengujian *Breusch-Pagan* dengan hipotesis pada taraf nyata pengujian 5% sebagai berikut:

$$H_0 : \sigma^2_1 = \sigma^2_2 = \dots = \sigma^2_{10} = \sigma^2 \text{ (variansi antar lokasi sama)}$$

$$H_1 : \sigma^2_i \neq \sigma^2 \text{ (variansi antar lokasi tidak sama/terdapat heterogenitas spasial)}$$

dengan  $i = 1, 2, \dots, 10$  dan  $\sigma^2 = \text{varian dari } y$

Tolak  $H_0$  jika nilai  $BP > \chi^2_{(\alpha,p)}$  atau  $p - \text{value} < \alpha$ .

**Tabel 5. Hasil pengujian *Breusch-Pagan***

<i>Breusch-Pagan</i>	<i>p-value</i>
8,0885	0,04422

Nilai *Breusch-Pagan* lebih besar dari  $\chi^2_{(0,05;3)} = 7.814728$ . Selain itu, nilai *p-value* yang lebih kecil dari taraf nyata pengujian 5% juga menunjukkan bahwa  $H_0$  ditolak yang artinya terdapat heterogenitas spasial.

#### 5. Pembobotan Spasial

Adanya heterogenitas spasial menyebabkan perlu adanya matriks pembobotan spasial. Fungsi pembobotan yang digunakan dalam penelitian ini adalah *adaptive bisquare kernel*. Langkah awal untuk membentuk matriks pembobot yaitu menghitung jarak *Euclid*. Selanjutnya menentukan lebar jendela (*bandwidth*), pemilihan *bandwidth* yang optimal menggunakan teknik *Golden Section Search* yang dilakukan secara iterasi menggunakan kriteria *Cross Validation (CV)*. Nilai CV yang minimum menunjukkan nilai *bandwidth* yang optimal [3]. Misalnya, untuk Kabupaten Bengkulu Selatan dengan *bandwidth* optimum 94,3112 km memiliki matriks pembobot  $W(u_1, v_1)$  yaitu sebagai berikut:

$$W(u_1, v_1) = [1; 0; 0; 0,726; 0,226; 0; 0; 0,003; 0; 0]$$

Jarak *Euclid* Kabupaten Bengkulu Selatan lebih besar dari nilai *bandwidth*-nya maka nilai pembobot kabupaten tersebut adalah nol, sedangkan jika jarak *Euclid* kurang dari nilai *bandwidth* akan diperoleh nilai yang bukan nol. Hal yang sama untuk matriks pembobot kabupaten/kota lainnya di Provinsi Bengkulu.

#### 6. Model Regresi Binomial Negatif Terboboti Geografis (RBNTG)

Model ini akan menghasilkan pendugaan parameter lokal dengan masing-masing lokasi akan memiliki parameter yang berbeda. Estimasi parameter dilakukan menggunakan metode Pendugaan Kemungkinan Maksimum *Likelihood* dengan memasukkan pembobot spasial dalam perhitungannya melalui iterasi *Newton Raphson*. Proses iterasi tersebut dilakukan pada setiap lokasi kabupaten/kota di provinsi Bengkulu menggunakan matriks pembobot spasial yang disesuaikan

masing-masing lokasi. Iterasi dihentikan ketika nilai estimasi parameter model sudah konvergen, yaitu apabila  $\hat{\alpha}^*_{(m+1)} \approx \hat{\beta}_{(m)}$  atau  $\|\hat{\beta}_{(m+1)} - \hat{\beta}_{(m)}\| \leq \varepsilon$  dimana  $\varepsilon$  nilai yang sangat kecil. Pada penelitian ini nilai  $\varepsilon = 10^{-3}$ .

**a. Uji kesamaan model Binomial Negatif dan RBNTG**

Hipotesis yang digunakan adalah sebagai berikut:

$H_0: \beta_k(u_i, v_i) = \beta_k$  (tidak terdapat perbedaan yang signifikan antara model Binomial Negatif dan model RBNTG)

$H_1: \beta_k(u_i, v_i) \neq \beta_k$ , dengan  $k = 1,2,3$  dan  $i = 1,2, \dots, 10$  (terdapat perbedaan yang signifikan antara model Binomial Negatif dan model RBNTG)

Daerah penolakannya adalah  $F_{hit} > F_{(\alpha, db_{RBNTG}, db_{BN})}$ .

**Tabel 6. Nilai hasil pengujian kesamaan model**

Model	Devians/db	$F_{Hitung}$	$F_{(0,05;6;6)}$
RBNTG	2,4999	1,2325	4,2838
BN	1,9975		

Berdasarkan Tabel 6 diperoleh bahwa nilai  $F_{Hitung}$  lebih kecil dari nilai  $F_{(0,05;6;6)}$  yaitu  $1,2325 < 4,2838$  sehingga  $H_0$  tidak ditolak, yang artinya dengan menggunakan taraf nyata pengujian 5% tidak terdapat perbedaan yang signifikan antara model Binomial Negatif dan model RBNTG.

**b. Pengujian Simultan Parameter Model RBNTG**

Hipotesis yang digunakan adalah sebagai berikut:

$H_0: \beta_1(u_i, v_i) = \beta_2(u_i, v_i) = \beta_3(u_i, v_i) = 0$  (semua variabel memberikan pengaruh yang signifikan terhadap model RBNTG)

$H_1$ : paling sedikit ada satu  $\beta_k(u_i, v_i) \neq 0$ , dengan  $k = 1,2,3$  dan  $i = 1,2, \dots, 10$  (paling sedikit ada satu variabel yang memberikan pengaruh signifikan terhadap model RBNTG)

Daerah penolakannya adalah  $D > \chi^2_{(\alpha,p)}$

**Tabel 7. Nilai Devians model RBNTG**

Devians	$\chi^2_{(0,05;3)}$
14,999	7.814

Tabel 4.11 menunjukkan bahwa nilai Devians lebih besar daripada nilai  $\chi^2_{(0,05;3)}$ , maka  $H_0$  ditolak. Artinya dengan menggunakan taraf nyata pengujian 5% paling sedikit ada satu variabel yang memberikan pengaruh signifikan terhadap model RBNTG dimana  $k = 1,2,3$  dan  $i = 1,2, \dots, 10$ .

**c. Pengujian Parsial Parameter Model RBNTG**

Hipotesis yang digunakan adalah sebagai berikut:

$H_0 : \beta_k(u_i, v_i) = 0$  (variabel  $k$  tidak memberikan pengaruh yang signifikan terhadap variabel respon pada tiap-tiap lokasi)

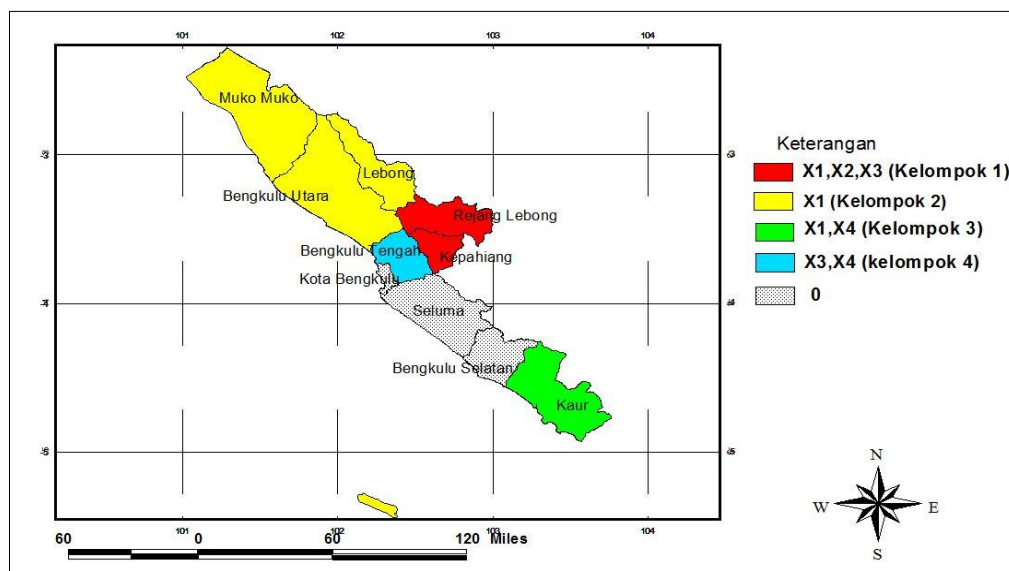
$H_1 : \beta_k(u_i, v_i) \neq 0$ , dengan  $k = 1,2,3$  dan  $i = 1,2, \dots, 10$  (variabel  $k$  memberikan pengaruh yang signifikan terhadap variabel respon pada tiap-tiap lokasi)

Daerah penolakannya adalah  $|Z_{hitung}| > Z_{\alpha/2}$ , dimana  $Z_{\alpha/2} = 1,96$ .

Berdasarkan hasil pengujian diperoleh bahwa variabel yang memberikan pengaruh signifikan terhadap jumlah kasus malaria berbeda-beda pada tiap lokasi kabupaten/kota di Provinsi Bengkulu, dari 10 lokasi hanya terdapat 7 lokasi yang memiliki variabel yang berpengaruh signifikan, sisanya 3 kabupaten/kota lainnya tidak terdapat variabel yang berpengaruh signifikan yaitu kabupaten Bengkulu Selatan, Seluma dan Kota Bengkulu. Sehingga kabupaten/kota di Provinsi Bengkulu dapat dikelompokkan menjadi 4 kelompok berdasarkan variabel yang berpengaruh signifikan dapat dilihat pada Tabel 8.

**Tabel 8. Pengelompokan kabupaten/kota berdasarkan variabel yang signifikan**

Kelompok	Kabupaten/kota	Variabel yang signifikan
1	Rejang Lebong, Kepahiang	$X_1, X_2, X_3$
2	Bengkulu Utara, Mukomuko, Lebong	$X_1$
3	Kaur	$X_1, X_3$
4	Bengkulu Tengah	$X_2, X_3$



**Gambar 1. Peta Pengelompokan kabupaten/kota berdasarkan variabel yang signifikan**

Gambar 1 menunjukkan bahwa pengelompokan yang terdiri dari 4 kelompok kabupaten/kota berdasarkan variabel yang signifikan tidak membentuk pola menyebar atau setiap kabupaten/kota

dalam kelompok tersebut letaknya saling berdekatan. Terlihat pada kelompok 1 dan kelompok 2 yang beranggotakan lebih dari satu kabupaten/kota yang saling berdekatan dan kelompok 3 dan 4 yang hanya terdiri dari satu kabupaten/kota. Terdapat kabupaten/kota yang tidak memiliki variabel yang berpengaruh signifikan terhadap jumlah kasus malaria yaitu kabupaten Seluma, Bengkulu Selatan dan Kota Bengkulu.

#### **4.1 Perbandingan Model Regresi Binomial Negatif dan Model RBNTG**

Salah satu kriteria yang digunakan untuk memilih model terbaik adalah *Akaike Information Criterion* (AIC), model terbaik antara dua model yang dibandingkan adalah model dengan nilai AIC terkecil [6]. Berdasarkan perhitungan yang telah dilakukan diperoleh bahwa nilai AIC pada Regresi Binomial Negatif 146,57 dan nilai AIC pada Regresi Binomial Negatif Terboboti Geografis 22,99. Sehingga dapat diketahui bahwa nilai AIC pada RBNTG lebih kecil dibandingkan AIC Regresi Binomial Negatif, maka dapat dikatakan bahwa Regresi Binomial Negatif Terboboti Geografis lebih baik digunakan untuk menganalisis jumlah kasus malaria sehingga dapat diketahui faktor-faktor yang mempengaruhi di setiap lokasi kabupaten/kota di Provinsi Bengkulu karena memiliki nilai AIC yang lebih kecil.

### **D. KESIMPULAN DAN SARAN**

#### **1. Kesimpulan**

Data jumlah kasus malaria di Provinsi Bengkulu merupakan data cacah yang mengalami overdispersi sehingga diperlukan regresi Binomial Negatif. Namun, adanya heterogenitas spasial pada data maka perlu Regresi Binomial Negatif Terboboti Geografis yang mampu untuk mengatasi overdispersi pada data cacah dengan memperhatikan efek spasial.

Faktor-faktor yang berpengaruh signifikan terhadap jumlah kasus malaria berbeda-beda pada tiap kabupaten/kota di Provinsi Bengkulu, dari 10 lokasi tersebut hanya terdapat 7 kabupaten/kota yang memiliki variabel yang berpengaruh signifikan, sisanya tidak terdapat variabel yang berpengaruh signifikan yaitu Kabupaten Bengkulu Selatan, Seluma dan Kota Bengkulu. Sehingga terdapat 4 kelompok berdasarkan variabel yang berpengaruh signifikan yang membentuk pola mengelompok atau berdekatan.

## DAFTAR PUSTAKA

- [1] BPS, 2014, *Indikator Kesejahteraan Rakyat Provinsi Bengkulu*, BPS Provinsi Bengkulu, Bengkulu.
- [2] Cameron, A.C., Trivedi, P.K., 1998, *Regression Analysis of Count Data*, Cambridge University Press, United Kingdom.
- [3] Cleveland, W.S., 1979, Robust Locally Weighted Regression and Smoothing Scatterplots, *Journal of America Statistics Association.*, 74, 829-836.
- [4] Li. F, (2000), *Multicollinearity*, Department of Statistics, Stockholm University.
- [5] McCullagh P dan Nelder J.A., 1989, *Generalized Linear Models Second Edition*, Chapman and Hall, London.
- [6] Nakaya, T., Fotheringham, A.S., Brunson, C. dan Charlton, M. 2005. Geographically Weighted Poisson Regression for Disease Association Mapping, *Statistics in Medicine*, Volume 24 Issue17, Pages 2695-2717.
- [7] Purhadi dan Ratri, N.H., 2015, Analisis Faktor-Faktor yang Mempengaruhi Jumlah Kasus Malaria di Jawa Timur Tahun 2013 dengan Geographically Weighted Negative Binomial Regression (GWNBR), *Jurnal Sains dan Seni ITS*, Volume 4 No 2, 2337-3520.
- [8] Silva A.R dan Rodrigues T.C.V., 2014, *Geographically Weighted Negative Binomial Regression-Incorporating Overdispersion*. Business Media New York, Springer Science.