

Model Log-Linier Beberapa Kasus Kriminologi Yang Terjadi di Wilayah Polres Bengkulu pada Tahun 2004/2005

Ely Mulyani¹⁾, Sigit Nugroho²⁾, Fachri Faisal²⁾

¹⁾Alumni Jurusan Matematika FMIPA Universitas Bengkulu

²⁾Staf Pengajar Jurusan Matematika FMIPA Universitas Bengkulu

ABSTRAK

Analisis Log-linier digunakan untuk menganalisis pola hubungan antara sekelompok variabel kategori yang mencakup asosiasi dua peubah, asosiasi tiga peubah atau lebih. Pola hubungan antar peubah dapat dilihat dari interaksi antar peubah-peubah itu sendiri. Tujuan dari penelitian ini adalah menunjukkan peubah-peubah yang cenderung dependen pada karakteristik pelaku tindak kriminal pencurian dan pencurian kendaraan bermotor yang terjadi di wilayah Polres Bengkulu berdasarkan model Log-linier yang terbentuk. Analisis data dilakukan dengan langkah-langkah: 1) Menyusun data ke dalam tabel kontingensi lima dimensi. 2) Menguji kesesuaian model secara simultan (goodnes of fit test), kesesuaian model diuji dengan membandingkan G^2 dari masing-masing model dengan $\chi^2_{(\alpha, db)}$ pada tingkat kepercayaan 95% ($\alpha = 0.05$). 3) Menguji kebebasan secara parsial, signifikansi hubungan yang ada diuji dengan membandingkan nilai G^2 parsial dengan $\chi^2_{(\alpha, db)}$ pada tingkat kepercayaan 95% ($\alpha = 0.05$). 4) Penyeleksian model terbaik dengan metode backward elimination.

Hasil penelitian menunjukkan: 1) Untuk pelaku tindak kriminal pencurian; adanya interaksi antara usia dengan pendidikan, usia dengan pekerjaan, usia dengan TKP, pendidikan dengan pekerjaan, dan pendidikan dengan TKP. 2) Untuk pelaku tindak kriminal pencurian kendaraan bermotor; adanya interaksi jenis kelamin dengan usia, pendidikan dengan pekerjaan, pekerjaan dengan TKP.

Kata Kunci: Model Log-linier, kriminal, data kategori, tabel kontingensi.

PENDAHULUAN

Kejahatan merupakan bentuk pelanggaran terhadap kaidah-kaidah sosial, jika kaidah-kaidah sosial yang oleh masyarakat itu dianggap sebagai nilai atau aturan yang dijunjung tinggi, mendapat pelanggaran dan penyimpangan dari orang atau sekelompok orang, maka hal itu dianggap sebagai gejala kejahatan karena dengan pelanggaran dan penyimpangan yang dilakukan itu, masyarakat merasa hidupnya menjadi terguncang atau terganggu (Abdulsyani, 1987).

Variabel yang menunjukkan karakteristik pelaku tindak kejahatan, pada umumnya berupa data kategori. Untuk mengetahui variabel-variabel kategorikal yang menimbulkan dependensi/keterkaitan, maka perlu kajian tentang Analisis Model Log-Linier (David, 1992). Dengan terbentuknya Model Log-Linier dapat diketahui faktor-faktor yang menimbulkan dependensi antar variabel yang berpengaruh pada pelaku tindak kriminal. Analisis log linear digunakan untuk menganalisis pola hubungan antar sekelompok peubah kategori yang mencakup asosiasi dua peubah, asosiasi tiga peubah atau lebih, baik secara simultan maupun secara parsial. Pola hubungan antar peubah dapat dilihat dari interaksi antara peubah itu sendiri.

Penelitian ini bertujuan untuk menunjukkan faktor-faktor yang dependen pada karakteristik pelaku tindak kriminal untuk tiap tindak kriminal yang diamati yang terjadi di wilayah Polresta Bengkulu berdasarkan Model Log-Linier yang terbentuk dan menentukan interaksi faktor-faktor yang dependen pada karakteristik pelaku tindak kriminal yang diamati berdasarkan Model Log-Linier yang terbentuk.

TINJAUAN PUSTAKA

Secara harfiah yang dimaksud dengan kriminologi adalah suatu ilmu yang khusus mempelajari hal ihwan kejahatan (crimen=aib), sedangkan yang dimaksud dengan poenologi adalah suatu ilmu yang mempelajari tentang penghukuman atau hal penjatuhan hukuman (poena=tebusan). Letak hubungan antara hukum pidana dengan kriminologi dan poenologi adalah dalam hal objeknya, yakni kejahatan dan penghukumannya yang dasar pandangannya adalah sebagai berikut: Menurut hukum pidana yang disebut kejahatan adalah segala tindak manusia yang melawan keharusan yang telah ditentukan berlaku sebagai hukum dan memenuhi perumusan ketentuan hukum pidana; Menurut kriminologi dan poenologi kejahatan ialah segala sikap tindak manusia yang menyeleweng, sehingga merupakan suatu gejala sosial yang menimbulkan reaksi negatif pula dari masyarakat (Soesilo, R. 1976).

Berhubung dengan ini, kriminologi biasanya dibagi menjadi tiga bagian: *Biology Criminal*, yang menyelidiki dalam diri orang itu akan sebab-sebab dari perbuatannya, baik dalam jasmani maupun rohaninya; *Sociology Criminal*, yang mencoba mencari sebab-sebab dalam lingkungan masyarakat di mana penjahat itu berada; *Policy Criminal*, yaitu tindakan-tindakan apa yang sekiranya harus dijalankan supaya orang lain tidak berbuat jahat (Bonger, 1982).

Faktor-faktor yang mempengaruhi seseorang melakukan kejahatan dapat berupa faktor-faktor *Biological* maupun *Psychological* dan faktor di luar diri pribadi pelaku, yaitu berupa lingkungan atau kelompok masyarakat di mana pelaku tersebut berada. Kedua faktor tersebut diperinci lagi menjadi beberapa faktor yaitu faktor keturunan, faktor penyakit jiwa (epilepsi, gejala sosiopatik, *schizophrenic*), faktor rumah tangga dan keluarga (*broken home*, penerapan disiplin dan latihan dalam keluarga), faktor ekonomi, serta faktor lingkungan (Chibro, 1984).

Analisis Model Log- Linier

Analisis model Log-Linear berguna untuk menentukan dependensi antara beberapa variabel kategorikal, kelebihan metode ini adalah (Agresti, 1990):

Dapat menentukan model matematika yang sesuai untuk dependensi multivariabel.

Dapat digunakan untuk menentukan interaksi yang menyebabkan variabel tersebut dependen.

Misalkan dalam sejumlah observasi ditentukan tiga peubah yakni 1, 2 dan 3, masing-masing peubah terdiri dari sejumlah kategori. Untuk mencari semua kemungkinan dengan mencari model secara teratur dan berurutan dari ordo yang lebih tinggi menuju ordo yang lebih rendah, menggunakan prinsip hirarki, yaitu jika ordo yang mempunyai tingkatan lebih tinggi masuk dalam model, maka ordo yang lebih rendah juga masuk dalam model, begitu sebaliknya jika ordo yang lebih rendah tidak masuk dalam model maka efek ordo yang lebih tinggi tidak masuk dalam model.

Kemungkinan model yang dihasilkan adalah (Anonim, 2005):

Model Independen

Taksiran nilai harapan untuk Model independen log linier adalah:

$$\hat{m}_{ijk} = \left(\frac{X_{i**}}{N} \right) \left(\frac{X_{*j*}}{N} \right) \left(\frac{X_{**k}}{N} \right) N$$

$$= \frac{(X_{i**})(X_{*j*})(X_{**k})}{N^2}$$

$$\log \hat{m}_{ijk} = \log X_{i**} + \log X_{*j*} + \log X_{**k} - 2 \log N$$

Maka bentuk umum model log linier independen adalah:

$$\log \hat{m}_{ijk} = \mu + \lambda 1_i + \lambda 2_j + \lambda 3_k$$

dengan : $\lambda 12_{ij} = \lambda 13_{ik} = \lambda 23_{jk} = \lambda 123_{ijk} = 0$; model ini hanya ada pengaruh rata-rata umum dan pengaruh utama peubah-peubah tanpa dipengaruhi oleh interaksi antar peubah.

Model Satu Interaksi Dua Faktor.

Taksiran nilai harapannya adalah:

$$\hat{m}_{ijk} = \left(\frac{X_{ij^*}}{N} \right) \left(\frac{X_{**k}}{N} \right) N$$

$$= \frac{(X_{ij^*})(X_{**k})}{N}$$

Bentuk model log linier satu interaksi dua faktor adalah

$$\log \hat{m}_{ijk} = \mu + \lambda 1_i + \lambda 2_j + \lambda 3_k + \lambda 12_{ij}$$

dengan: $\lambda 13_{ik} = \lambda 23_{jk} = \lambda 123_{ijk} = 0$; model ini menyatakan dependensi antara variabel 1 dan variabel 2, dengan variabel 3 ada, atau signifikansi dalam model.

begitu juga kemungkinan model log linier yang menyatakan dependensi antara variabel 1 dan variabel 3, dengan variabel 2 ada, atau signifikan dalam model. Juga model log linier yang menyatakan dependensi antara variabel 2 dan variabel 3, dengan variabel 1 ada, atau signifikan dalam model dapat dinyatakan seperti prosedur di atas.

Model Dua Interaksi Dua Faktor

Taksiran nilai harapannya adalah

$$\hat{m}_{ijk} = \frac{(X_{ij^*})(X_{i^*k})}{X_{i^{**}}}$$

Bentuk model log linier dua interaksi dua faktor adalah

$$\text{Log } \hat{m}_{ijk} = \mu + \lambda 1_{(i)} + \lambda 2_{2(j)} + \lambda 3_{(k)} + \lambda 12_{(ij)} + \lambda 13_{(ik)}$$

dengan: $\lambda 23_{jk} = \lambda 123_{ijk} = 0$; model ini menyatakan adanya dependensi antara variabel 1 dan 2, dan variabel 1 dan 3.

kemungkinan model log linier yang menyatakan dependensi antara variabel 1 dan 2, dan variabel 2 dan 3; Juga model yang menyatakan dependensi antara variabel 1 dan variabel 3, dan variabel 2 dengan variabel 3 dapat dinyatakan dengan cara yang sama seperti di atas.

Model Tanpa Interaksi Tiga Faktor

Taksiran nilai harapannya adalah

$$\hat{m}_{ijk} = \frac{(X_{ij^*})(X_{i^*k})(X_{*jk})}{(X_{i^{**}})(X_{*j^*})(X_{**k})}$$

Bentuk model log linier tanpa interaksi tiga faktor adalah

$$\widehat{\text{Log } m_{ijk}} = \mu + \lambda 1_{(i)} + \lambda 2_{2(j)} + \lambda 3_{(k)} + \lambda 12_{(ij)} + \lambda 13_{(ik)} + \lambda 23_{(jk)}$$

dengan: $\lambda 123_{ijk} = 0$; model ini semua interaksi dua faktor ada atau signifikan dalam model, tetapi tidak ada interaksi atau dependensi antara ketiga faktor.

Model Saturated

Model Saturated adalah model yang memuat semua parameter yang mungkin dan tidak dapat dimasuki parameter-parameter lainnya (Isgianto, 2004).

Modelnya adalah:

$$\widehat{\text{Log } m_{ijk}} = \mu + \lambda 1_{(i)} + \lambda 2_{(j)} + \lambda 3_{(k)} + \lambda 12_{(ij)} + \lambda 13_{(ik)} + \lambda 23_{(jk)} + \lambda 123_{(ijk)}$$

Dengan kendala-kendala (constraints):

$$\sum_{i=1}^I \lambda 1_{(i)} = \dots = \sum_{i=1}^I \sum_{j=1}^J \lambda 12_{(ij)} = \dots = \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K \lambda 123_{(ijk)} = 0$$

Keterangan:

m_{ijk} = frekuensi diharapkan pada setiap sel ke-ijk dalam model

X_{i**} = menyatakan jumlah observasi pada baris ke-i

X_{*j*} = menyatakan jumlah observasi pada kolom ke-j

X_{***k} = menyatakan jumlah observasi pada lapisan ke-k

μ = pengaruh rata-rata umum

$1_{(i)}$ = pengaruh utama faktor 1 kategori ke-i

$2_{(j)}$ = pengaruh utama faktor 2 kategori ke-j

$3_{(k)}$ = pengaruh utama faktor 3 kategori ke-k

$12_{(ij)}$ = pengaruh interaksi 1 dan 2 kategori ke-i

$13_{(ik)}$ = pengaruh interaksi 1 dan 3 kategori ke-ik

$23_{(jk)}$ = pengaruh interaksi 2 dan 3 kategori ke-jk

$123_{(ijk)}$ = pengaruh interaksi 1, 2 dan 3 kategori ke-ijk

Prosedur Log-Linier

Uji Kesesuaian Model Secara Simultan

Uji kesesuaian model secara simultan dilakukan dengan menggunakan k-faktor. Dalam pengujian ini terdapat dua (2) langkah pengujian pemodelan, yaitu:

Uji kesesuaian model dengan pengaruh dari interaksi k-faktor dan yang lebih tinggi sama dengan nol.

Uji ini bertujuan untuk mengetahui jumlah minimal peubah yang berintegrasi di dalam model log-linier terbaik. Kesesuaian model diuji dengan membandingkan G^2 dari masing-masing model dengan

χ^2_{tabel} pada tingkat kepercayaan tertentu. Statistik likelihood ratio-nya adalah:

$$G^2 = 2 \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K n_{ijk} \log \left[\frac{n_{ijk}}{m_{ijk}} \right]$$

dimana:

n_{ijk} = banyaknya pengamatan pada sel ke- ijk
 m_{ijk} = banyaknya nilai harapan pada sel ke- ijk
Pengujian Hipotesis dilakukan sebagai berikut:
Hipotesis yang diuji:

H_0 = interaksi k-faktor dan yang lebih tinggi = 0 efek order ke-3 atau lebih = 0

H_1 = interaksi k-faktor dan yang lebih tinggi $\neq 0$ (nyata ada dalam model)

Statistik uji: digunakan tes χ^2 untuk k independen.

Tingkat kepercayaan: digunakan $\alpha = 0,05$

Daerah penolakan: daerah penolakan terdiri dari semua nilai G yang besar, sehingga kemungkinan yang berkaitan dengan terjadinya nilai-nilai tersebut sama atau lebih kecil dari $\alpha = 0,05$. Hipotesis nol (H_0) ditolak apabila nilai uji statistik $\chi^2_{observasi} (G^2) > \chi^2_{tabel}$.

Apabila hipotesis alternatif diterima, tiap model memberi arti sebagai berikut:

k=3; menunjukkan bahwa ada hubungan antara 3 peubah secara simultan.

k=2; menunjukkan bahwa ada hubungan antara 2 peubah dan 3 peubah secara simultan.

k=1; menunjukkan bahwa ada hubungan antara 1 peubah, 2 peubah dan 3 peubah secara simultan.

Uji kesesuaian model dengan pengaruh dari interaksi k-faktor sama dengan nol.

Untuk memastikan tes k-faktor pertama dan melihat interaksi mana yang nilainya sama dengan nol, maka dilakukan tes pada interaksi k-faktor sama dengan nol, kesesuaian model diuji dengan membandingkan G dari masing-masing model dengan χ^2_{tabel} pada tingkat kepercayaan tertentu.

Pengujian hipotesis dilakukan sebagai berikut:

Hipotesis yang diuji:

H_0 = interaksi k-faktor = 0.

H_1 = interaksi k-faktor $\neq 0$ (nyata ada dalam model)

Statistik uji: digunakan tes χ^2 untuk k independen.

Tingkat kepercayaan: digunakan $\alpha = 0,05$

Daerah penolakan: daerah penolakan terdiri dari semua nilai G yang besar, sehingga kemungkinan yang berkaitan dengan terjadinya nilai-nilai tersebut sama atau lebih kecil dari $\alpha = 0,05$. Hipotesis nol (H_0) ditolak apabila nilai uji statistik $\chi^2_{observasi} (G^2) > \chi^2_{tabel}$

Apabila hipotesis alternatif diterima, tiap model memberi arti sebagai berikut:

k=1; menunjukkan bahwa ada efek utama dari masing-masing peubah.

k=2; menunjukkan bahwa ada hubungan antara 2 peubah secara simultan.

k=3; menunjukkan bahwa ada hubungan antara 3 peubah secara simultan.

Uji Kebebasan Secara Parsial.

Pengujian ini bertujuan untuk mengetahui efek utama dan interaksi antar peubah mana yang secara nyata ada dalam model. Misalkan pada uji secara simultan diketahui ada efek utama dan interaksi antar peubah yang nyata dalam model, dalam uji ini dimungkinkan hanya ada beberapa efek utama dan interaksi antarpeubah yang nyata dalam model.

Signifikansi hubungan yang ada diuji dengan membandingkan nilai G^2 parsial dengan χ^2_{tabel} pada tingkat kepercayaan tertentu.

Pengujian hipotesis dilakukan sebagai berikut:

Hipotesis yang diuji:

H_0 = tak ada interaksi antar berbagai peubah.

H_1 = ada interaksi antar berbagai peubah.

Statistik uji: digunakan tes χ^2 untuk k independen.

Tingkat kepercayaan: digunakan $\alpha = 0,05$

Daerah penolakan: daerah penolakan terdiri dari semua nilai G yang besar, sehingga kemungkinan yang berkaitan dengan terjadinya nilai-nilai tersebut sama atau lebih kecil dari $\alpha = 0,05$. Hipotesis nol (H_0) ditolak apabila nilai uji statistik $\chi^2_{observasi} (G^2) > \chi^2_{tabel}$

Penyeleksian Model Terbaik.

Menyeleksi model terbaik dengan metode elimination backward dimulai dari model penuh (saturated) dan secara berurutan mengeliminasi model. Pada setiap tahap, metode ini mengeliminasi model yang mempunyai pengaruh/efek yang paling kecil. Proses seleksi berhenti ketika beberapa pengeliminasian yang dilakukan telah mendapatkan model yang sesuai dan lebih sederhana.

Hipotesis yang digunakan adalah sebagai berikut:

H_0 = model (2) adalah model yang terbaik

H_1 = model (1) adalah model yang terbaik

Statistik uji yang digunakan: $G^2_{(2-1)} = G^2_{(2)} - G^2_{(1)}$

Dengan: $G^2_{(2)}$ = statistik likelihood G^2 untuk model (2)

$G^2_{(1)}$ = statistik likelihood G^2 untuk model (1)

METODOLOGI PENELITIAN

Penelitian ini merupakan penelitian terapan (applied research), yang menjadi populasi dalam penelitian ini adalah seluruh pelaku tindak kejahatan yang diamati. Tindak kejahatan yang diamati itu adalah pencurian dengan kekerasan, dan pencurian ken daraan bermotor dari Bulan Juni 2004- Juli 2005. Total pelaku tindak kejahatan tersebut adalah sebanyak 126 orang. Sampel yang digunakan adalah total sampling, yaitu keseluruhan anggota populasi dijadikan sebagai sampel.

Analisis model Log-linier untuk mendapatkan model terbaik menggunakan metode backward dengan bantuan software SPSS (HIERARCHICAL LOG LINIER).

Tabel 1. Daftar Peubah dan kategori yang digunakan.

| Nama Peubah | Peubah | Kategori |
|-------------------------------|--------|------------------------|
| Jenis kelamin | A | 1. laki-laki |
| | | 2. perempuan |
| Usia | B | 1. anak-anak |
| | | 2. remaja |
| | | 3. dewasa |
| Pendidikan | C | 1. SD |
| | | 2. SMP |
| | | 3. SMA |
| Pekerjaan | D | 1. PNS/TNI/POLRI |
| | | 2. wiraswasta/karyawan |
| | | 3. pekerja lepas |
| | | 4. tidak bekerja |
| Tempat Kejadian Perkara (TKP) | E | 1. Gadaing Cempaka |
| | | 2. Muara Bangkahulu |
| | | 3. Teluk Segara |
| | | 4. Selebar |
| | | 5. Kampung Melayu |
| | | 6. Ratu Agung |
| | | 7. Ratu Samban |
| | | 8. Sungai Serut |

HASIL DAN PEMBAHASAN

Dari hasil penelitian, untuk pelaku tindak kriminal pencurian terlihat bahwa: dari metode backward elimination dihasilkan model log linier terbaik adalah:

$[C * D] [B * C] [B * D] [B * E] [C * E] [A]$. Dari model log linier terbaik yang dihasilkan dengan menggunakan metode backward elimination di atas dapat ditulis model Log Linier yang berdasarkan prinsip hierarki, yang berarti jika interaksi yang lebih kompleks masuk dalam model maka interaksi yang lebih sederhana juga akan masuk dalam model. Begitu juga sebaliknya, jika interaksi yang lebih sederhana tidak masuk dalam model maka interaksi yang lebih kompleks tidak masuk dalam model.

Dalam persamaan log linier ditulis sebagai berikut:

$$\log m_{ijklm} = \mu + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_l^D + \lambda_m^E + \lambda_{jk}^{BC} + \lambda_{jl}^{BD} + \lambda_{jm}^{BE} + \lambda_{kl}^{CD} + \lambda_{km}^{CE}$$

Model di atas menunjukkan adanya interaksi antara :Usia dengan pendidikan, usia dengan pekerjaan, usia dengan tempat kejadian perkara (TKP), pendidikan dengan pekerjaan, dan pendidikan dengan TKP.

Untuk pelaku tindak kriminal pencurian kejahatan bermotor, dengan model backward elimination dihasilkan model Log Linier terbaik adalah:

$[C * D] [A * B] [D * E]$

Dalam persamaan model Log Linier ditulis sebagai berikut:

$$\log m_{ijklm} = \mu + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_l^D + \lambda_m^E + \lambda_{ij}^{AB} + \lambda_{kl}^{CD} + \lambda_{lm}^{DE}$$

Model di atas menunjukkan adanya interaksi antara :jenis kelamin dengan usia, pendidikan dengan pekerjaan, pekerjaan dengan TKP.

KESIMPULAN

Dalam penelitian ini data pelaku tindak kejahatan berupa data kategori, sehingga dapat digunakan analisis Log Linier, yang berfungsi untuk menganalisis pola hubungan antara sekelompok variabel kategori yang mencakup asosiasi dua peubah, asosiasi tiga peubah atau lebih. Untuk pelaku tindak kriminal pencurian terlihat adanya interaksi antara:usia dengan pendidikan, usia dengan pekerjaan, usia dengan tempat kejadian perkara (TKP), pendidikan dengan pekerjaan, dan pendidikan dengan TKP. Untuk pelaku tindak kriminal pencurian kejahatan bermotor terlihat adanya interaksi jenis kelamin dengan usia, pendidikan dengan pekerjaan, pekerjaan dengan TKP. Tentu hasil ini bukan merupakan suatu hal yang mutlak karena beberapa faktor luaran lain seperti cara pengambilan sampel dan teknik survey yang dilakukan mungkin mempengaruhi hasil yang didapat.

DAFTAR PUSTAKA

- [1]. Agresti, A.1990. Categorical Data Analysis. New York : John Wiley & Sons.
- [2]. Agung, IGN. 2001. STATISTIKA ; Penerapan Metode Analisis Untuk Tabulasi Sempurna dan Tak Sempurna. Jakarta : PT Raja Grafindo Persada.
- [3]. Agung, IGN. 2002. STATISTIKA ; Analisis Hubungan Kausal Berdasarkan Data Kategorik. Jakarta : PT Raja Grafindo Persada.
- [4]. Anonim, 2004. http://www.cas.lancs.ac/courses/inter_SPSS.htm.
- [5]. Anonim, 2004. Log-Linier Model Analysis in SAS and SPSS. <http://www.uky.edu/computingcenter>
- [6]. Anonim, 2005. <http://www.statsoftinc.com/textbook/stloglin.htm>.
- [7]. Arikunto, S. 2002. Prosedur Penelitian : Suatu pendekatan Praktek. Jakarta: Rineka Cipta.
- [8]. Bandung : Alummni.

- [9]. Bonger. 1987. Pengantar Tentang Kriminologi. Jakarta : Ghalia Indonesia.
- [10]. David, C. H. 1992. Statistical Methods for Psychology. Third Edition. California : Duxbury Press.
- [11]. Everitt, B.S. 1977. The Analysis of Contingency Tables. John Wiley & Sons, Inc. New York, New York, USA.
- [12]. Hamzah, A. 1986. Bunga Rampai Hukum Pidana dan Acara Pidana. Jakarta : Ghalia Indonesia.
- [13]. Hosmer, D. W. and S.Lemeshow. 1989. Applied Logistic Regression. New York : John Wiley & Sons.
- [14]. Isgianto, A. 2004. Pemodelan Dependensi Data Kategori Melalui Pendekatan Model Log Linier. Bengkulu : Hibah SP4 Kompetisi Universitas Bengkulu Tahun I.
- [15]. Moeljatno. 1987. Azas - Azas Hukum Pidana. Jakarta : P.T Bina Aksara
- [16]. Purbacaka, P and A. Ridwan Halim. 1997. Filsafat Hukum Pidana ; dalam tanya jawab. Jakarta : C.V Rajawali.
- [17]. Simanjuntak dan Ali Chaidir. 1978. Cakrawala baru Kriminologi. Bandung : Tarsito.
- [18]. Soejono, D. 1978. Penanggulangan Kejahatan (Crime Prevention). Cetakan II.
- [19]. Soejono, D. 1978. Pengantar Kriminologi. Bandung : P.T Tribisana Karya.
- [20]. Soesilo, R. 1976. Kriminologi (Pengetahuan Tentang Sebab - Sebab Kejahatan). Bogor : Politea.

Penggunaan Metode Jumlah Kuadrat Galat Terkecil Terboboti Untuk Mengeliminasi Pengaruh Sumber Keragaman Yang Tidak Dapat Dikendalikan

Lila Sulastri¹, Sigit Nugroho², dan Fachri Faisal²

1 Alumni Jurusan Matematika FMIPA Universitas Bengkulu

2 Staf Pengajar Jurusan Matematika FMIPA Universitas Bengkulu

ABSTRAK

Fluktuasi hasil yang diakibatkan oleh pengaruh sumber keragaman yang tidak dapat dikendalikan khususnya suhu, cuaca dan iklim merupakan masalah yang sering dijumpai dalam penelitian bidang pertanian terutama pada penelitian yang berlangsung dalam jangka waktu yang lama. Hal ini mempersulit upaya pengungkapan informasi seperti pengaruh faktor perlakuan, tendensi dan stabilitas hasil tanaman berkaitan dengan varietas dan taraf suatu perlakuan. Penggunaan metode kuadrat terkecil terboboti dalam mengeliminasi pengaruh sumber keragaman yang tidak dapat dikendalikan khususnya suhu, cuaca dan iklim memberikan suatu kemungkinan untuk mengeliminasi fluktuasi hasil tersebut. Gagasan dasarnya adalah mentransformasi peubah Y menjadi peubah lain Z. Hal yang penting dalam metode kuadrat terkecil terboboti dalam mengeliminasi pengaruh sumber keragaman yang tidak dapat dikendalikan tersebut adalah bagaimana pembobot (w_i) dapat ditentukan berdasarkan data pengamatan. Disamping akan mempermudah penelusuran informasi yang diperlukan untuk penelitian suatu varietas unggul komoditi pertanian, metode ini juga dapat digunakan sebagai alternatif transformasi data pada analisis ragam untuk mendapatkan sebaran yang memenuhi asumsi yang disyaratkan.

Kata kunci: Fluktuasi hasil, pembobot, stabilitas hasil, metode kuadrat terkecil terboboti

PENDAHULUAN

Hal yang penting dalam metode ini adalah bagaimana pembobot (w_i) dapat ditentukan berdasarkan data pengamatan, kemudian menggandakan w_i yang bersesuaian dengan model yang diasumsikan, dan menerapkan prinsip metode jumlah kuadrat galat terkecil yaitu meminimumkan jumlah kuadrat simpangan terboboti $\sum w_i \varepsilon_i^2$.

Dalam tulisan ini, metode jumlah kuadrat galat terkecil terboboti diaplikasikan pada suatu varietas tanaman yang mengalami fluktuasi hasil pada beberapa kali musim tanam. Fluktuasi produksi per masa tanam yang disebabkan oleh perubahan suhu, cuaca dan iklim dapat dieliminasi dengan menggunakan metode jumlah kuadrat galat terkecil terboboti. Dengan cara ini diharapkan pengaruh faktor-faktor pertumbuhan dan faktor lainnya yang merupakan faktor penting dalam penelitian pengembangan kualitas varietas tanaman yang lebih unggul dalam kondisi lingkungan manapun dapat dievaluasi secara lebih baik. Validitas yang tinggi dan informasi penting lainnya juga dapat diperoleh melalui metode ini. Disamping akan mempermudah penelusuran informasi yang diperlukan, metode ini juga dapat digunakan sebagai alternatif transformasi data pada analisis ragam untuk mendapatkan sebaran data yang memenuhi asumsi yang disyaratkan (Suharjo, 1997).

MASALAH DAN TUJUAN

Pelanggaran-pelanggaran terhadap asumsi jumlah kuadrat galat terkecil terutama keheteroskedastisitasan yang terjadi pada data penelitian dapat diteliti dengan menggunakan plot sisaan. Untuk menerapkan metode jumlah kuadrat galat terkecil terboboti perlu ditentukan pembobot yang bersesuaian kemudian mentransformasi peubah agar memenuhi asumsi dasar yang disyaratkan. Dalam aplikasinya terhadap kasus fluktuasi hasil tanaman, metode ini digunakan untuk mengeliminasi pengaruh sumber keragaman yang tidak dapat dikendalikan khususnya suhu, cuaca dan iklim yang menyebabkan terjadinya fluktuasi hasil tersebut.

Oleh karena itu perlu diketahui bagaimana metode jumlah kuadrat galat terkecil terboboti digunakan untuk menduga parameter dalam menentukan model yang optimal sehingga pengaruh sumber keragaman yang tidak dapat dikendalikan dapat dieliminasi. Kemudian dapat dibandingkan seberapa besar keakuratan metode jumlah kuadrat galat terkecil terboboti dari metode jumlah kuadrat galat terkecil biasa dalam menduga parameter untuk suatu model regresi jika data mengalami heteroskedastisitas.

Hasil penelitian ini dapat dijadikan sebagai masukan bagi penelitian bidang pertanian dalam mengeliminasi sumber keragaman tak terkendalik pada lahan pertanaman suatu komoditi pertanian yang mengalami fluktuasi hasil produksi. Jika tidak dilakukan analisis dengan menggunakan pembobot pada data asli maka dugaan yang didapat tidak akan memberikan gambaran stabilitas hasil produksi (Suharjo, 1997). Apabila stabilitas hasil produksi tidak diketahui, sukar meneliti faktor penelitian untuk pengembangan kualitas varietas tanaman tersebut.

Asumsi

Adapun yang menjadi asumsi dalam penelitian ini adalah (Weisberg, 1985) :

Y_i saling bebas satu sama lain.

ε_i adalah variabel acak dan mean ε_i adalah nol

$E(\varepsilon_i) = 0, i = 1, 2, \dots, n.$

varian (ragam) penduga $V(\varepsilon_i) = \sigma^2 < \infty, i = 1, 2, \dots, n.$

galat tidak berkorelasi, yaitu $Cov(\varepsilon_i, \varepsilon_j) = 0, i \neq j.$

$\varepsilon_i \sim N(0, \sigma^2), i = 1, 2, \dots, n.$

Metode Kuadrat Terkecil Terboboti

Pembobot

Pembobot diberikan agar ditemukan model yang baru yang memenuhi asumsi Homoskedastisitas. Sehingga pada model tersebut dapat diterapkan hal-hal yang bersangkutan dengan metode kuadrat terkecil. Secara umum pembobot itu merupakan kebalikan dari ragamnya ($w_i = 1/\sigma_i^2$) (Steel and Torrie, 1981). Pada dasarnya pembobot itu dibatasi pada data yang memiliki ulangan dan σ_i^2 dapat diduga melalui ragam setiap taraf peubah bebas. Dengan mentransformasikan amatan Y menjadi peubah Z yang memenuhi asumsi dasar maka dapat digunakan prosedur kuadrat terkecil terboboti.

Misalkan secara umum model yang digunakan adalah $Y = X\beta + \varepsilon$

$$E(\varepsilon) = 0, V(\varepsilon) = V\sigma^2, \text{ dan } \varepsilon \sim N(0, V\sigma^2)$$

Akan diperlihatkan bahwa ada suatu matriks simetris P yang tidak singular yang bersifat $P'P = PP = P^2 = V$.

Misalkan bahwa $\varepsilon^* = P^{-1}\varepsilon$, sehingga $E(\varepsilon^*) = 0$.

Selanjutnya akan dibuktikan bahwa jika ε^* adalah suatu peubah acak vektor dengan $E(\varepsilon^*) = 0$, maka $E(\varepsilon^*(\varepsilon^*)') = V(\varepsilon^*)$ dengan operasi nilai harapan diterapkan pada setiap unsur matrik $\varepsilon^*(\varepsilon^*)'$ yang berukuran $n \times n$ itu. Dengan demikian

$$\begin{aligned} V(\varepsilon^*) &= E(\varepsilon^*(\varepsilon^*)') = E(P^{-1}\varepsilon\varepsilon'P^{-1}), \text{ karena } (P^{-1})' = P^{-1} \\ &= P^{-1}E(\varepsilon\varepsilon')P^{-1} \\ &= P^{-1}PPP^{-1}\sigma^2 \\ &= I\sigma^2 \end{aligned}$$

Juga dapat ditunjukkan bahwa $\varepsilon^* \sim N(0, I\sigma^2)$, yang berarti ε^* menyebar normal karena unsur-unsur ε^* merupakan kombinasi linier unsur-unsur ε yang diasumsikan menyebar normal.

Bila $\mathbf{Y} = \mathbf{X}\beta + \varepsilon$ diganda awalkan dengan \mathbf{P}^{-1} maka diperoleh model baru $\mathbf{P}^{-1}\mathbf{Y} = \mathbf{P}^{-1}\mathbf{X}\beta + \mathbf{P}^{-1}\varepsilon$ atau bisa juga ditulis sebagai

$$\mathbf{Z} = \mathbf{Q}\beta + \varepsilon^* \quad (1)$$

Karena $E(\varepsilon^*) = 0$ dan $\mathbf{V}(\varepsilon^*) = \mathbf{I}\sigma^2$ maka metode kuadrat terkecil dapat diterapkan pada persamaan $\mathbf{Z} = \mathbf{Q}\beta + \varepsilon^*$. Jumlah kuadrat sisanya adalah

$$(\varepsilon^*)'\varepsilon^* = \varepsilon'\mathbf{V}^{-1}\varepsilon = (\mathbf{Y} - \mathbf{X}\beta)'\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\beta).$$

Persamaan normal $\mathbf{Q}'\mathbf{Q}\mathbf{b} = \mathbf{Q}'\mathbf{Z}$ menjadi

$\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}$ dengan solusi

$$\mathbf{b} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y} \quad (2)$$

bila matrik yang dibalik itu tidak singular. Jumlah kuadrat regresinya adalah $\mathbf{b}'\mathbf{Q}'\mathbf{Z} = \mathbf{Y}'\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}$ sedangkan jumlah kuadrat totalnya adalah $\mathbf{Z}'\mathbf{Z} = \mathbf{Y}'\mathbf{V}^{-1}\mathbf{Y}$.

Jumlah kuadrat galat dapat dinyatakan se bagai $[\mathbf{Z}'\mathbf{Z}] - [\mathbf{b}'\mathbf{Q}'\mathbf{Z}]$. Jumlah kuadrat yang berasal dari nilai tengah adalah $(\sum Z_i)^2/n$, dengan Z_i adalah unsur vektor \mathbf{Z} . Matrik varian-covarian bagi \mathbf{b} adalah

$$\mathbf{V}(\mathbf{b}) = (\mathbf{Q}'\mathbf{Q})^{-1}\sigma^2 = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\sigma^2 \quad (3)$$

Sisaan yang harus diperiksa adalah nilai dugaan bagi $\varepsilon^* = \mathbf{P}^{-1}\varepsilon$. Sisaan-sisaan itu dapat diperoleh menurut rumus $\mathbf{P}^{-1}(\mathbf{Y} - \hat{\mathbf{Y}})$ dengan $\hat{\mathbf{Y}} = \mathbf{X}\mathbf{b}$ dan \mathbf{b} diperoleh dari persamaan (2). Dengan demikian sisaan tersebut

$$\mathbf{P}^{-1}\{\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\}\mathbf{Y} \quad (4)$$

Penerapan paling sederhana metode kuadrat terkecil terboboti adalah bila amatan –amatan bebas satu sama lain namun mempunyai ragam yang berbeda sehingga

$$\mathbf{V}\sigma^2 = \begin{bmatrix} \sigma_1^2 & & & 0 \\ & \sigma_2^2 & & \\ & & \ddots & \\ 0 & & & \sigma_n^2 \end{bmatrix}$$

Bila kuadrat terkecil biasa yang digunakan, walaupun seharusnya kuadrat terkecil terboboti, maka nilai-nilai dugaannya diperoleh dari $\mathbf{b}_0 = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ dan

$$E(\mathbf{b}_0) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\beta = \beta \quad (5)$$

namun

$$\begin{aligned} \mathbf{V}(\mathbf{b}) &= \mathbf{V}((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}) \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}(\mathbf{Y})\mathbf{X}'(\mathbf{X}'\mathbf{X})^{-1} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}'(\mathbf{X}'\mathbf{X})^{-1}\sigma^2 \end{aligned} \quad (6)$$

Jika analisis yang dilakukan benar maka persamaan (3) terpenuhi dan secara umum, unsur-unsur matriks ini akan menghasilkan ragam yang lebih kecil bagi masing-masing koefisien maupun bagi fungsi linier koefisien.

Sebagai ilustrasi sederhana misalkan terdapat model $E(Y) = \beta X$

Misalkan pula

$$\mathbf{V}(\mathbf{Y}) = \mathbf{V}\sigma^2 = \begin{bmatrix} 1/w_1 & & & 0 \\ & 1/w_2 & & \\ & & \ddots & \\ 0 & & & 1/w_n \end{bmatrix}$$

Dari matriks \mathbf{V} akan ditemukan

$$V^{-1} = \begin{bmatrix} w_1 & & & 0 \\ & w_2 & & \\ & & \ddots & \\ 0 & & & w_n \end{bmatrix}$$

Dengan menerapkan hasil-hasil umum di atas, maka dapat diperoleh

$$b = \frac{\sum w_i X_i Y_i}{\sum w_i X_i^2}$$

dengan semua penjumlahan dilakukan untuk semua $i = 1, 2, \dots, n$

Kasus 1. Misalkan bahwa $\sigma_i^2 = V(Y_i) = kX_i$, artinya ragam setiap Y_i sebanding dengan besarnya nilai X_i . Maka $w_i = \sigma^2 / kX_i$. Dengan demikian

$$b = \frac{\sum Y_i}{\sum X_i}$$

Jadi, jika ragam setiap Y_i sebanding dengan nilai X_i , maka nilai dugaan terbaik bagi koefisien regresi dalam model tersebut adalah rataan Y_i dibagi dengan rataan X_i . Selain itu $\mathbf{v}(\mathbf{b}) = \frac{k}{\sum X_i}$

Kasus 2. Misalkan bahwa $\sigma_i^2 = V(Y_i) = kX_i^2$, artinya ragam setiap Y_i sebanding dengan kuadrat nilai X_i padananya, maka $w_i = \sigma^2 / kX_i^2$. Dengan demikian

$$b = \frac{\sum \frac{Y_i}{X_i}}{n}$$

Jadi, jika ragam setiap Y_i sebanding dengan X_i^2 , maka nilai dugaan terbaik bagi koefisien regresi adalah rata-rata n kemiringan (slope) yang diperoleh dari setiap pasang amatan Y_i/X_i . Juga $\mathbf{v}(\mathbf{b}) = \frac{k}{n}$

Meminimumkan jumlah kuadrat simpangan $\sum (\varepsilon_i^*)^2$, pada model juga akan dilakukan dalam penentuan nilai dugaan bagi parameter regresi. Untuk setiap pengamatan, Jumlah Kuadrat Simpangan persamaan (1) memenuhi $\sum (\varepsilon_i^*)^2 = \sum (Z_i - \beta_0^* + \beta_1^* M_i)^2$

Peubah Z tersebut dinyatakan kembali ke dalam peubah Y , sehingga

$$\begin{aligned} \sum (\sqrt{w_i} \varepsilon_i)^2 &= \sum (\sqrt{w_i} Y_i - \beta_0 \sqrt{w_i} - \beta_1 X_i \sqrt{w_i})^2 \\ \sum w_i \varepsilon_i^2 &= \sum w_i (Y_i - \beta_0 - \beta_1 X_i)^2 \end{aligned} \quad (7)$$

Pendugaan parameter

Dimisalkan nilai dugaan bagi parameter tersebut adalah b_0^* dan b_1^* , hal ini untuk membedakan dengan nilai dugaan kuadrat terkecil biasa.

$$b_1^* = \frac{\sum w_i (x_i - \bar{x}^*) (y_i - \bar{y}^*)}{\sum w_i (x_i - \bar{x}^*)^2} \quad (8)$$

dan

$$b_0^* = \frac{\sum w_i y_i - b_1^* \sum w_i x_i}{\sum w_i} \quad (9)$$

dengan \bar{y}^* dan \bar{x}^* adalah rata-rata terboboti untuk masing-masing peubah.

Kemudian akan ditunjukkan bahwa sifat-sifat penduga dalam kasus heteroskedastisitas dengan metode kuadrat terkecil terboboti yaitu penduga tersebut merupakan kombinasi linier dari peubah acak Y , tak bias, dan memiliki ragam minimum.

Residual

Residual dalam analisis regresi terboboti adalah simpangan antara nilai amatan terboboti dengan nilai regresi terboboti. Nilai yang bersesuaian adalah $w_i Y_i$ dan $w_i \hat{Y}_i$ sehingga sisaan (ϵ_i) adalah $w_i (Y_i - \hat{Y}_i)$. Dengan menerapkan uji residual dapat dilihat kelayakan model dengan memplot residualnya.

Kelayakan Model Regresi Terboboti

Seperti pada metode kuadrat terkecil biasa untuk mengukur kelayakan, dilakukan salah satunya uji F. Hal ini dimaksudkan untuk melihat apakah model regresi yang didapatkan telah layak atau tidak. Jumlah kuadrat yang digunakan dalam metode alternatif ini adalah jumlah kuadrat total terboboti, jumlah kuadrat regresi terboboti dan jumlah kuadrat residu terboboti dengan rumus sebagai berikut :

$$JK \text{ Total} = \sum w_i (y_i - \bar{y})^2 \tag{11}$$

$$JK \text{ Regresi} = \frac{\left\{ \sum w_i y_i x_i - \left[\left(\sum w_i y_i \right) \left(\sum w_i x_i \right) / \sum w_i \right] \right\}^2}{\sum w_i x_i - \left(\sum w_i x_i \right)^2 / \sum w_i} \tag{12}$$

$$JK \text{ Residual} = JK \text{ Total} - JK \text{ Regresi} \tag{13}$$

Dengan adanya JK tersebut dapat ditentukan kelayakan sebuah regresi dan dapat dilakukan berbagai pengujian terhadap persamaan regresi yang diperoleh seperti uji F, uji ketakpasan model, koefisien determinasi, dan pengujian residual. Semua pengujian tersebut memiliki cara yang sama seperti yang dilakukan dalam metode kuadrat terkecil biasa.

Eliminasi Pengaruh Faktor Tak Terkendalkan dengan Metode Kuadrat Terkecil Terboboti

Di bidang pemuliaan tanaman, adalah umum bila penelitian dilakukan pada lokasi yang berbeda dalam kurun waktu tertentu untuk mendapatkan informasi mengenai pengaruh berbagai kondisi lingkungan (Suharjo, 1997). Untuk mengatasi munculnya fluktuasi hasil tanaman baik disebabkan oleh perbedaan lokasi maupun beragamnya keadaan lingkungan dan faktor iklim digunakan nilai rata-rata per lokasi sebagai pembobot data asli untuk mendapatkan nilai dugaan (Fox and Rosielle, 1982)

Stabilitas hasil merupakan suatu informasi penting dalam pemuliaan tanaman. Besaran ini mencerminkan kemampuan produksi suatu varietas tanaman pada berbagai kondisi lingkungan. Konsep stabilitas yang populer saat ini yaitu varietas tanaman dikatakan stabil bila (Becker dalam Suharjo, 1997)

- keragaman hasilnya pada lingkungan rendah (biological concept).
- interaksi dengan kondisi lingkungannya rendah (agronomic concept).

Untuk mengeliminasi munculnya fluktuasi hasil yang disebabkan oleh faktor lingkungan khususnya suhu, cuaca dan iklim digunakan metode eliminasi yang dilandaskan pada pembobotan masing-masing hasil tanaman dengan nilai rata-rata per lokasi setiap masa tanam.

Cara yang digunakan untuk mengetahui manfaat eliminasi pengaruh iklim melalui metode ini, adalah melalui perbandingan setiap hasil dengan data asli terhadap konsep stabilitas. Selain itu juga akan dibandingkan analisis ragam masing-masing jenis data tersebut.

Hilangnya fluktuasi hasil akibat sumber keragaman yang tidak dapat dikendalikan dalam beberapa musim tanam terhadap komoditi yang diteliti dengan menerapkan metode kuadrat terkecil terboboti akan memudahkan penelusuran terhadap kestabilan taraf-taraf perlakuan yang sedang diteliti, karena hasil yang diberikan melalui berbagai kriteria kestabilan menunjukkan adanya kekonsistenan (Suharjo, 1997).

METODE PENELITIAN

Adapun yang menjadi populasi dalam penelitian ini adalah tanaman Jagung Varietas unggul Bisi 2 pada lahan percobaan di wilayah Riau Kepulauan dengan 5 taraf percobaan lokasi lahan yang luasnya berbeda pada lima kali masa tanam. Metode jumlah kuadrat galat terkecil terboboti ini diaplikasikan dalam pendugaan parameter untuk model regresi yang sesuai bagi kasus fluktuasi hasil tahunan tanaman Jagung Bisi 2 di wilayah Riau kepulauan. Pada prakteknya tidak ada lahan yang memiliki kondisi yang homogen (Suharjo, 1997), perbedaan dapat muncul dari berbagai aspek, baik yang dapat dikendalikan maupun yang tidak dapat dikendalikan. Pada daerah yang sama dalam kurun waktu yang berbeda kondisi iklim sebagai sumber keragaman yang tidak dapat dikendalikan dapat sangat beragam. Untuk mendapatkan informasi mengenai pengaruh faktor pertumbuhan yang tak terkendali perlu dilakukan penelitian pada beberapa musim tanam pada lebih dari satu lokasi, sehingga perilaku tanaman pada berbagai kondisi iklim dan lingkungan yang berbeda dapat dievaluasi. (Fox and Rosielle, 1982).

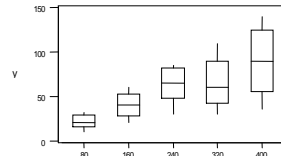
HASIL DAN PEMBAHASAN

Tabel 1. Data pengujian bibit jagung bisi 2 pada beberapa musim tanam dan luas lahan yang berbeda

| Masa tanam | luas lahan (m ²) | hasil produksi (kg) |
|-------------------------|------------------------------|---------------------|
| Mei-Agustus 1997 | 80 | 25 |
| September-Desember 1997 | 160 | 35 |
| Januari-April 1998 | 240 | 65 |
| Mei-Agustus 1998 | 320 | 70 |
| September-Desember 1998 | 400 | 75 |
| Januari-April 1999 | 80 | 10 |
| Mei-Agustus 1999 | 160 | 20 |
| September-Desember 1999 | 240 | 65 |
| Januaru-April 2000 | 320 | 55 |
| Mei-Agustus 2000 | 400 | 35 |
| September-Desember 2000 | 80 | 20 |
| Januari-April 2001 | 160 | 40 |
| Mei-Agustus 2001 | 240 | 30 |
| September-Desember 2001 | 320 | 35 |
| Januari-April 2002 | 400 | 90 |
| Mei-Agustus 2002 | 80 | 20 |
| September-Desember 2002 | 160 | 45 |
| Januari-April 2003 | 240 | 80 |
| Mei-Agustus 2003 | 320 | 60 |
| September-Desember 2003 | 400 | 110 |
| Januari-April 2004 | 80 | 32 |
| Mei-Agustus 2004 | 160 | 60 |
| September-Desember 2004 | 240 | 85 |
| Januari-April 2005 | 320 | 110 |
| Mei-Agustus 2005 | 400 | 140 |

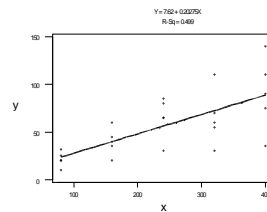
Terjadinya fluktuasi hasil tanam ini mengakibatkan pengaruh faktor-faktor percobaan terhadap hasil tanaman yang diteliti menjadi sukar dievaluasi, karena keragaman hasil yang berasal dari faktor perlakuan terkacaukan oleh keragaman akibat pengaruh iklim (Suharjo, 1997). Hal ini secara

eksploratif dapat dilihat pada box plot (Gambar 1) yang menggambarkan sebaran pengamatan masing-masing taraf perlakuan.

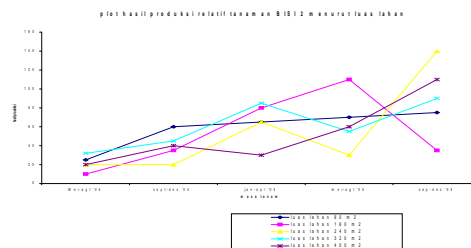


Gambar 1. Box Plot hasil produksi menurut luas lahan

Tendensi hasil yang merupakan salah satu informasi penting yang berkaitan dengan penelitian faktor - faktor keragaman terkendalikan, selama kurun waktu penelitian disajikan pada gambar 2 dan 3. Dapat dilihat bahwa melalui pendekatan pola hubungan linier antara hasil tanaman dan waktu tanam terlihat simpangan yang sangat besar dari waktu ke waktu. Regresi linier antara hasil dan faktor penelitian berupa luas lahan juga menunjukkan adanya simpangan yang cukup besar terhadap garis linier.



Gambar 2. Plot hasil produksi terhadap faktor luas lahan

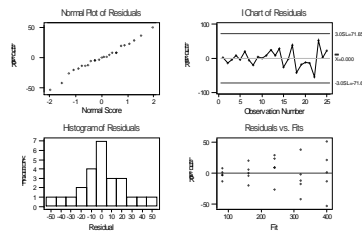


Gambar 3. Plot tendensi hasil produksi tanaman Jagung per musim tanam

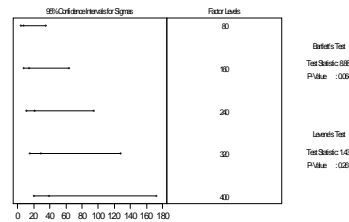
Fluktuasi demikian besar, meskipun pola umum masih dapat dilihat, namun simpangan tiap -tiap pengamatan terhadap garis regresi linier cukup besar. Hal ini mengakibatkan informasi yang dijelaskan oleh garis regresi secara analisis tertutupi oleh simpangan pengamatan tersebut.

Penelitian terhadap sumber keragaman yang dapat dikendalikan, misalnya faktor -faktor genetik tanaman yang harus diteliti demi penemuan varietas unggul yang tahan terhadap berbagai kondisi lingkungan akan sukar dilakukan dalam keadaan kondisi hasil yang mengalami fluktuasi yang cukup besar. Oleh karena itu untuk menemukan suatu kestabilan diperlukan adanya pengeliminasian terhadap pengaruh sumber keragaman yang tidak dapat dikendalikan khususnya iklim. Pengeliminasian ini pada prinsipnya adalah menemukan model yang mampu menunjukkan stabilitas hasil dalam berbagai kondisi (Suharjo, 1997). Pendugaan parameter dalam kasus ini tidak cukup layak dengan menggunakan metode kuadrat terkecil biasa karena adanya pelanggaran asumsi yang disyaratkan yaitu heteroskedastisitas. Keheteroskedastisitasan data dapat dilihat pada Gambar 4 dan

hasil uji kehomogenan ragam dapat dilihat pada Gambar 5. Kedua plot menunjukkan heteroskedastisitas yang sangat mencolok.

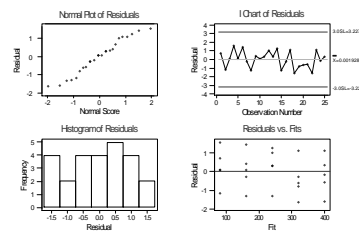


Gambar 4. Plot residu tanaman Jagung Bisi 2

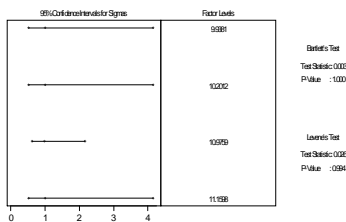


Gambar 5. Plot tes kehomogenan varian

Keheteroskedastisitasan mudah diamati melalui plot residual. Dengan melihat plot residual dari data awal terbukti bahwa terjadi kemenjuluran (skewness) pada scatter plot residual. Untuk melihat sejauh mana model yang didapatkan melalui analisis regresi terboboti yang meminimumkan varians, dilakukan kembali pemplotan terhadap residual dari model $y = 0.243 x$.



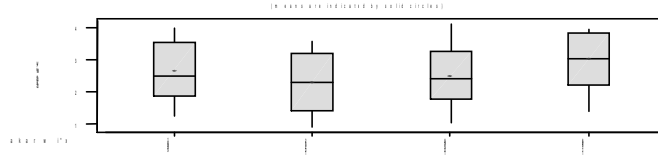
Gambar 6. Plot residual regresi terboboti



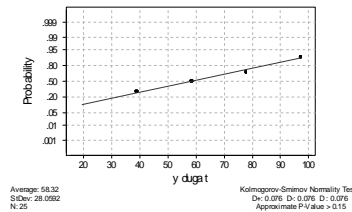
Gambar 7. Plot tes kehomogenan varian data terboboti

Pola plot residual yang ditunjukkan pada Gambar 6 memperlihatkan pola tebaran residual yang tidak lagi mengalami skewness (kemenjuluran) meskipun belum menggambarkan tebaran residual yang memuaskan secara sempurna namun dapat diartikan memuaskan. Plot test kehomogenan ragam data yang telah diboboti pada Gambar 7 juga mendukung bahwa keheterogenan ragam dapat diatasi. Teratasinya ketidakstabilan keragaman hasil yang berasal dari faktor perlakuan yang pada awalnya

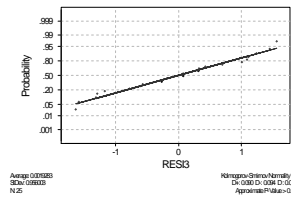
terkacaukan oleh keragaman akibat pengaruh cuaca dan iklim, secara eksploratif dapat dilihat pada box plot (Gambar 8)



Gambar 8. Box plot hasil tanaman menurut jenis lahan setelah diboboti. Melihat terpenuhi atau tidaknya asumsi kenormalan dapat digunakan plot berikut ini.



Gambar 9. Normal Plot



Gambar 10. Normal plot residual

Pola plot diatas memperlihatkan terpenuhinya asumsi kenormalan, sehingga tidak ada alasan untuk mengatakan data menyebar tidak normal. Selain itu pemeriksaan terhadap residual juga menunjukkan bahwa residual menyebar normal. Hal ini dapat dilihat dari normal plot residual, disamping itu didukung oleh statistik Kolmogorooov Smirnov. Dari Gambar 10 dapat dilihat bahwa nilai D uji Kolmogorooov Smirnov (0.094) lebih kecil dari nilai D tabel pada taraf 0.15 (0.220) yang berarti residual berdistribusi normal.

Pendugaan parameter yang dilakukan dengan metode kuadrat terkecil terboboti untuk kasus fluktuasi hasil pada tanaman Jagung varietas unggul Bisi 2 yang telah dilakukan ini menunjukkan bahwa semua asumsi yang disyaratkan sebagai penduga yang baik telah terpenuhi. Berdasarkan ketetapan awal dari penelitian bahwa suatu penduga yang memenuhi semua asumsi yang disyaratkan hanya akan didapatkan dengan menggunakan analisis yang benar (Draper and Smith, 1992), maka disimpulkan analisis regresi terboboti merupakan salah satu analisis yang benar untuk kasus fluktuasi hasil tanaman ini dalam mengeliminasi sumber keragaman tak dapat dikendalikan.

Koeffisien Determinasi untuk model yang telah didapatkan melalui analisis regresi terboboti dalam penelitian fluktuasi hasil tanaman ini adalah :

$$R^2 = \frac{\sum(\hat{Y}_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2} = \frac{166.89}{188.47} = 88.6\%$$

Jadi, persamaan $y = 0.243 x$ untuk kasus fluktuasi hasil tanaman ini dapat menjelaskan 88.6% dari keragaman total dalam data.

Ketepatan Model dan Galat Murni

Galat murni akan timbul jika data mengalami ulangan (replikasi), sebelumnya telah dinyatakan bahwa ulangan yang dimaksud adalah ulangan sesungguhnya. Seandainya bukan ulangan sesungguhnya s_e^2 cenderung menduga kurang (underestimate) σ^2 , dan uji ketepatan model cenderung mendeteksi secara salah ketidakpasan yang sesungguhnya tidak ada (Draper and Smith, 1992).

Analisis Terboboti dengan pembobot w_i melewati titik pusat
 Persamaan regresinya adalah $Y = 0.243 x$

| Prediktor | Koefisien | StDev | t | Nilai-p |
|----------------|-----------|---------|------|---------|
| X = Luas lahan | 0.24298 | 0.01784 | 3.62 | 0.000 |

Analisis Keragaman

| Sumber | db | JK | KT | F | Nilai-p |
|------------|----|--------|--------|--------|---------|
| Regression | 1 | 166.89 | 166.89 | 185.56 | 0.000 |
| Error | 24 | 21.58 | 0.90 | | |
| Total | 25 | 188.47 | | | |

Durbin-Watson statistic = 2.25
 Pure error test - F = 0.46 P = 0.7608
 DF(pure error) = 20

Dalam tabel analisis ragam tertera bahwa Pure error test - F = 0.46. Hal ini menunjukkan bahwa nilai F tidak nyata. Jadi berdasarkan uji ini tidak ada alasan untuk meragukan kepasan model dan menggunakan $s^2 = 0.9$ sebagai suatu dugaan bagi σ^2 untuk mengadakan uji F bagi regresi keseluruhan.

Perbandingan Keakuratan Kuadrat terkecil Terboboti dibandingkan dengan Kuadrat Terkecil Biasa. Kelayakan suatu model regresi biasanya diukur dengan menggunakan koefisien determinasi. Model ini memiliki nilai koefisien determinasi 49.9%. Sedangkan model regresi yang parameternya diduga dengan kuadrat terkecil terboboti memiliki koefisien determinasi 88.6%. Secara intuitif dapat dinilai bahwa analisis regresi terboboti memberikan model yang lebih valid jika dibandingkan dengan analisis kuadrat terkecil biasa.

Stabilitas Hasil Setelah Eliminasi Pengaruh Sumber Keragaman tak Terkendalikan dengan Analisis Regresi Terboboti. Salah satu syarat (asumsi) yang harus dipenuhi untuk mendapatkan hasil dengan validitas tinggi adalah dipenuhi kehomogenan ragam. Bila syarat ini dilanggar, maka pengujian hipotesis pada analisis ini tak lagi sensitif (Eisenhart dalam Suharjo, 1997). Penggunaan data asli dalam memberikan gambaran stabilitas hasil tidak memberikan validitas yang memuaskan terhadap konsep stabilitas, yaitu keragaman hasil pada berbagai kondisi lingkungan rendah (minimum) (biological concept). Keadaan ini dapat diatasi dengan menggunakan kuadrat terkecil terboboti dalam menduga parameter untuk model regresi fluktuasi hasil tanaman Jagung ini. Buktinya konsep stabilitas hasil dalam biological concept (Becker dalam Suharjo, 1997) dapat terpenuhi dengan menggunakan metode ini. Manfaat yang dapat diambil oleh para peneliti bidang pertanian dengan didapatkannya model regresi yang mampu memberikan gambaran stabilitas hasil pada berbagai kondisi lingkungan ini adalah memudahkan penelusuran informasi terhadap faktor-faktor penelitian yang mungkin ingin diketahui dalam pengembangan varietas baru yang lebih unggul (Pearce dalam Suharjo, 1997).

KESIMPULAN

Penggunaan kuadrat terkecil terboboti dalam estimasi parameter pada kasus tanaman Jagung yang mengalami fluktuasi hasil dalam penelitian bidang pertanian, memberikan beberapa keuntungan. Memudahkan penelusuran informasi penelitian terhadap sumber keragaman yang dapat dikendalikan terhadap tanaman dalam menemukan varietas unggul. Walaupun pada awalnya penggunaan model regresi dengan menggunakan data asli tidak memiliki validitas yang tinggi namun dengan menggunakan kuadrat terkecil terboboti dalam pendugaan parameter untuk sebuah model yang diinginkan dapat diatasi. Dengan analisis ini, pengaruh sumber keragaman yang tidak dapat dikendalikan yang menyebabkan model dari data asli tidak valid bisa dieliminasi sehingga

perbandingan antar taraf perlakuan menjadi mungkin dan data tetap memiliki sifat asli. Disamping itu tendensi hasil juga akan mudah dikenali. Hilangnya fluktuasi hasil antar masa tanam memudahkan penelusuran akan kestabilan suatu taraf perlakuan, karena hasil yang diberikan menunjukkan kekonsistenan.

DAFTAR PUSTAKA

- [1]. Chatterjee and Price. 1977. Regression Analysis by Example. USA: John Wiley and Sons Inc.
- [2]. Dixon, W.J. dan F. J. Massey Jr..1991. Pengantar Analisis Statistik (Terjemahan). Yogyakarta: Gadjahmada Universitas Press.
- [3]. Draper, N and H. Smith. 1992. Analisis Regresi Terapan (Edisi kedua) (Terjemahan). Jakarta: Gramedia Pustaaaka Utama.
- [4]. Dunn and Clark. 1974. Applied Statistics: Analysis of Variance and Regression. USA : John Wiley and Sons Inc.
- [5]. Foster, B. 2001. Fisika SMU. Bandung : Ganesha Operation.
- [6]. Fox, P.N. and A.A. Rosielle. 1982. Reducing the influence of environmental main effects on pattern analysis of plant breeding environments. Euphytica.
- [7]. Freud, J.E and R. E. Walpole. 1987. Mathematical Statistics 4th. Prentice Hall: Englewood Cliffe Inc.
- [8]. Gujarati, D. 1991. Ekonometrika Dasar. Jakarta: Erlangga.
- [9]. Herrhyanto, N. 2003. Statistika Matematis Lanjutan. Bandung: Pustaka Setia.
- [10]. James and James. 1976. Mathematics Dic tionary 4th. USA: Van Nostrand Reinhold.
- [11]. Kutner, M.H., C. J. Nachtsheim and friends. 2005. Applied Linier Statistical Models. Singapore: Mc. Graw-Hills
- [12]. Notoatmojo S. 2002. Metodologi Penelitian Kesehatan. Jakarta : Rineka Cipta.
- [13]. Soelistyo, 2001. Dasar-dasar Ekonometrika. Yogyakarta: BPFE
- [14]. Steel, R.G.D. and J. H. Torrie. 1981. Principles and Procedures Of Statistics A Biometrical Approach. Singapore: Mc Graw-Hills
- [15]. Steel, R.G.D., J. H. Torrie and David A. 1997. Principles and Procedures Of Statistics A Biometrical Approach 3th edition. New York: Mc Graw-Hills
- [16]. Suharjo, B. 1997. Eliminasi Pengaruh Iklim Pada Komoditi Pertanian Dalam Penelitian Jangka Panjang. Forum Statistika Dan Komputasi IPB. Vol.2, 36 -45.
- [17]. Walpole, R.E. 1992. Pengantar Statistika (Terjemahan). Jakarta: Gramedia Pustaka Utama.
- [18]. Weisberg, S. 1985. Applied Linier Regression. New York : John Wiley and Sons.
- [19]. Wonnacot and Wonnacot. 1981. Regression : Second Course In Statistics. New York: John Wiley and Sons Inc.

Analisis Komponen Utama Penerimaan Pajak dan Retribusi Kota Bengkulu

Trisea Oktaviana¹⁾, Sigit Nugroho²⁾, Fachri Faisal²⁾.

¹⁾Alumni Jurusan Matematika FMIPA Universitas Bengkulu

²⁾Staf Pengajar Jurusan Matematika FMIPA Universitas Bengkulu

ABSTRAK

Tujuan penelitian ini adalah untuk mengetahui sumber-sumber pendapatan asli daerah Kota Bengkulu yang membentuk komponen utama. Jenis Penelitian menggunakan penelitian terapan dengan metode analisis komponen utama. Data yang digunakan data sekunder dari kantor Dinas Pendapatan Daerah Kota Bengkulu. Data direduksi dengan menggunakan matriks ragam peragam dan matriks korelasi. Hasil penelitian menunjukkan bahwa matriks ragam peragam menghasilkan satu komponen utama yang menjelaskan keragaman peubah asal sebesar 99.801%. Dari matriks korelasi komponen yang digunakan empat komponen utama, yang menjelaskan keragaman sebesar 83.83%.

Kata Kunci : Komponen utama, matriks korelasi, matriks ragam peragam, dan pendapatan asli daerah.

PENDAHULUAN

Sejak Januari 2001 bangsa dan negara Indonesia melaksanakan otonomi daerah diseluruh daerah tingkat II. Kota Bengkulu merupakan daerah otonom, oleh karena itu perlu diberikan sumber - sumber penerimaan yang akan digunakan untuk menjalankan otonominya secara baik [2]. Untuk meningkatkan penerimaan pendapatan asli daerah dikota Bengkulu, perlu diketahui terlebih dahulu pendapatan asli daerah yang sumber-sumbernya berpengaruh besar terhadap penerimaan tersebut. Karena sumber-sumber tersebut sangat banyak, maka untuk mengetahuinya per lu dilakukan reduksi terhadap sumber-sumber tersebut. Penggunaan komponen utama, yang merupakan fungsi linier tertentu dari peubah asal sering disarankan untuk digunakan dalam proses mereduksi banyaknya peubah. Prosedur statistis untuk mendapatkan komponen utama yang mampu mempertahankan sebagian besar informasi yang terkandung pada data asal disebut sebagai analisis komponen utama [3]. Besarnya dana sumber-sumber pendapatan asli daerah senantiasa berubah setiap bulannya. Hal ini mempengaruhi besarnya penerimaan pendapatan asli daerah dikota Bengkulu. Tujuan penelitian ini adalah untuk mengetahui sumber-sumber pendapatan asli daerah yang sumbernya membentuk komponen utama penentu besarnya penerimaan pendapatan asli daerah kota Bengkulu.

TINJAUAN PUSTAKA

Seperti yang telah dikemukakan pada pendahuluan agar daerah otonom dapat melaksanakan fungsinya dengan baik, maka daerah otonom perlu diberikan sumber -sumber penerimaan daerah. Dengan adanya sumber pendapatan itu diharapkan daerah dapat menjalankan otonomin ya secara nyata dan tanggung jawab [13]. Pendapatan asli daerah adalah penerimaan daerah yang didapat dari usaha pemerintah daerah untuk mengumpulkan dana guna keperluan daerah dalam membiayai pengeluaran rutin maupun pembangunannya tiap tahun [5]. Analisis komponen utama adalah salah satu teknik eksplorasi data yang menghasilkan komponen utama. Dengan keragaman total yang didapat dari sedikit komponen utama mampu mempertahankan sebagian besar informasi yang terkandung pada data asal [3].

Misalkan notasi x_{jk} untuk menandai nilai tertentu dari pengamatan ke -j pada variabel ke-k maka n pengukuran pada p variabel dapat ditulis dalam bentuk matriks segi empat, misalkan matriks X dengan n baris dan p kolom [8].

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1k} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2k} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots & \dots & \vdots \\ x_{j1} & x_{j2} & \dots & x_{jk} & \dots & x_{jp} \\ \vdots & \vdots & \dots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nk} & \dots & x_{np} \end{bmatrix} \quad (1)$$

Rata-rata populasi secara umum dapat dihitung dari n pengukuran pada masing-masing p variabel, yaitu:

$$E(\mathbf{X}) = \begin{bmatrix} E(X_1) \\ E(X_2) \\ \dots \\ E(X_p) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \dots \\ \mu_p \end{bmatrix} = \boldsymbol{\mu} \quad (2)$$

Matriks ragam peragamnya adalah

$$\Sigma = cov(\mathbf{X}) = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2p} \\ \dots & \dots & \dots & \dots \\ \sigma_{p1} & \sigma_{p2} & \dots & \sigma_{pp} \end{bmatrix} \quad (3)$$

dan matriks korelasinya adalah

$$= \begin{bmatrix} \frac{11}{\sqrt{11}\sqrt{11}} & \frac{12}{\sqrt{11}\sqrt{22}} & \dots & \frac{1p}{\sqrt{11}\sqrt{pp}} \\ \frac{12}{\sqrt{11}\sqrt{12}} & \frac{22}{\sqrt{22}\sqrt{22}} & \dots & \frac{2p}{\sqrt{22}\sqrt{pp}} \\ \dots & \dots & \dots & \dots \\ \frac{1p}{\sqrt{11}\sqrt{pp}} & \frac{2p}{\sqrt{22}\sqrt{pp}} & \dots & \frac{pp}{\sqrt{pp}\sqrt{pp}} \end{bmatrix} = \begin{bmatrix} I & 12 & \dots & 1p \\ 12 & I & \dots & 2p \\ \dots & \dots & \dots & \dots \\ 1p & 2p & \dots & I \end{bmatrix} \quad (4)$$

Jika diberikan simpangan baku matriks p x p adalah

$$\mathbf{V}^{1/2} = \begin{bmatrix} \sqrt{11} & 0 & \dots & 0 \\ 0 & \sqrt{22} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sqrt{pp} \end{bmatrix} \text{ maka} \quad (5)$$

$$\mathbf{V}^{1/2} \mathbf{V}^{1/2} = \Sigma \text{ dan } (\mathbf{V}^{1/2})^{-1} \Sigma (\mathbf{V}^{1/2})^{-1}$$

Jika A adalah matriks n x n, maka vektor tak nol x didalam R^n dinamakan vektor ciri dari A. Jika Ax adalah kelipatan skalar dari x yakni, Ax = untuk suatu skalar, maka skalar dinamakan akar ciri dari A dan x dikatakan vektor ciri yang bersesuaian dengan.

Secara umum, komponen utama memiliki ciri sebagai berikut:

- Komponen utama merupakan kombinasi linier dari peubah asal.
- Tidak ada korelasi antar kompone n
- Mempunyai keragaman berurut dari yang terbesar ke yang terkecil. (Misalkan peubah asal X_1, X_2, \dots, X_p memiliki akar ciri $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$)

Kombinasi liniernya adalah

$$\begin{aligned} Y_1 &= a_1 \mathbf{X} = a_{11} X_1 + \dots + a_{1p} X_p \\ Y_2 &= a_2 \mathbf{X} = a_{21} X_1 + \dots + a_{2p} X_p \\ &\vdots \\ Y_p &= a_p \mathbf{X} = a_{p1} X_1 + \dots + a_{pp} X_p \end{aligned} \quad (6)$$

dengan $\sum a_i^2 = a_i' \sum a_i$ dan $cov(Y_i, Y_k) = cov(a_i \mathbf{X}, a_k \mathbf{X}) = a_i' \sum a_k = 0; i, k = 1, 2, \dots, p.$
dimana a_{ik} = vektor ciri dan Y_i = komponen utama ke-i.

Komponen utama adalah kombinasi linier Y_1, Y_2, \dots, Y_p dengan nilai ragam sebesar mungkin, yang berarti $\text{Var}(Y_1) \text{Var}(Y_2) \dots \text{Var}(Y_p)$, sehingga masing-masing komponen dapat dijelaskan sebagai berikut:

Komponen utama ke-1 = a_1X yang memaksimumkan $\text{Var}(a_1X)$ dimana $a_1' a_1 = 1$

Komponen utama ke-2 = a_2X yang memaksimumkan $\text{Var}(a_2X)$ dimana $a_2' a_2 = 1$ dan $\text{cov}(Y_1, Y_2) = \text{cov}(a_1X, a_2X) = 0$

Dan seterusnya hingga Komponen utama ke-i yaitu a_iX yang memaksimumkan $\text{Var}(a_iX)$ dimana $a_i' a_i = 1$ dan $\text{cov}(a_iX, a_kX) = 0$ untuk $k < i$.

Jika menggunakan matriks ragam peragam V , maka kontribusi dari setiap komponen utama ke-j, $j = 1, 2, \dots, p$ terhadap total keragaman X adalah

$$Y_j = \frac{\lambda_j}{\sum_{i=1}^p \lambda_i} \quad (7)$$

Dan korelasi antara peubah X_i dengan komponen utama ke-j, Y_j adalah

$$\text{corr}(X_i, Y_j) = a_{ji} \sqrt{\frac{\lambda_j}{\text{var}(X_i)}} \quad (8)$$

Serta skor komponen utama dari pengamatan ke -i, diperoleh dengan cara

$$\begin{aligned} KU_{1i} &= l_1'(x_i - \bar{x}) \\ KU_{2i} &= l_2'(x_i - \bar{x}) \quad \text{untuk } i = 1, \dots, n. \\ &\dots \\ KU_{pi} &= l_p'(x_i - \bar{x}) \end{aligned} \quad (9)$$

Besaran KU_{1i} merupakan skor komponen utama ke-1 dari pengamatan ke-i, KU_{ji} adalah skor komponen utama ke-j ($j = 1, 2, \dots, p$) dari pengamatan ke-i.

Matriks korelasi ρ digunakan jika p peubah asal yang diamati tidak semuanya menggunakan satuan pengukuran yang sama, maka peubah asal itu perlu dibakukan. Kontribusi komponen utama ke -j terhadap total keragaman adalah

$$Y_j = \frac{\lambda_j}{p} \quad (10)$$

dan korelasi antara peubah asal X_i dengan komponen utama ke-j, Y_j adalah

$$\text{corr}(X_i, Y_j) = a_{ji} \sqrt{\lambda_j} \quad (11)$$

Serta skor komponen utama dari pengamatan ke -i, didapat dari persamaan

$$\begin{aligned} KU_{1i} &= l_1' V^{-1/2} (x_i - \bar{x}) \\ KU_{2i} &= l_2' V^{-1/2} (x_i - \bar{x}) \\ &\dots \\ KU_{pi} &= l_p' V^{-1/2} (x_i - \bar{x}) \end{aligned} \quad \text{dengan} \quad V^{1/2} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ \sqrt{s_{11}} & & & \\ 0 & \frac{1}{\sqrt{s_{22}}} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \frac{1}{\sqrt{s_{pp}}} \end{bmatrix} \quad (12)$$

dan s_{ii} adalah unsur diagonal ke-i dari matriks [17].

Ada tiga metode yang umum digunakan untuk menentukan banyaknya komponen utama. Metode pertama didasarkan pada *kumulatif proporsi keragaman* total yang mampu dijelaskan. Pada metode ini, pemilihan komponen utama dikatakan ideal bila kumulatif proporsi keragamannya sebesar 80% - 90% [8]. Jika $\lambda_1 \lambda_2 \dots \lambda_p$ adalah akar ciri dari matriks ragam peragam maka proporsi kumulatif dari k komponen utama pertama adalah

$$\frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^p \lambda_i}, k = 1, \dots, p \quad (13)$$

Pada kasus penggunaan matriks korelasi maka $\sum_{i=1}^p \lambda_i = p$ sehingga proporsi kumulatifnya adalah

$$\frac{1}{p} \sum_{i=1}^k \lambda_i, k = 1, \dots, p \quad [3]. \quad (14)$$

Metode kedua hanya bisa diterapkan pada penggunaan matriks korelasi. Metode ini disarankan oleh *Kaiser* (1960) yang berargumen bahwa jika peubah asal saling bebas maka komponen utama tidak lain adalah peubah asal, dan setiap komponen utama akan memiliki ragam satu. Sehingga jika ada komponen utama yang ragamnya kurang dari satu dianggap memiliki kontribusi yang kurang [3]. Metode ketiga adalah penggunaan *grafik* yang disebut scree plot Ide yang ada dibelakang metode ini adalah pada scree yang mulai merata. Interpretasi terhadap plot ini sangat subjektif[3].

METODE PENELITIAN

Jenis penelitian ini merupakan penelitian terapan (applied research), yaitu mengkaji suatu metode dan mencoba diterapkan pada suatu kasus. Adapun metode yang digunakan adalah Analisis Komponen Utama. Data yang digunakan adalah data sekunder yang diperoleh dari Dinas Pendapatan Daerah Kota Bengkulu selama 2 tahun terakhir (2004 -2005). Variabel yang akan diteliti adalah sumber-sumber pendapatan asli daerah kota bengkulu yaitu pajak dan retribusi. Tahap-tahap analisis data dalam penelitian ini adalah sebagai berikut:

Mereduksi data dalam bentuk matriks korelasi dan matriks ragam peragam.

Mencari nilai akar ciri dan vektor ciri.

Menentukan Banyaknya Komponen Utama

Penentuan berdasarkan pada kumulatif proporsi total keragaman.

Penentuan berdasarkan pada ragam komponen utama (akar ciri).

Penentuan berdasarkan Scree Plot

Mencari Hubungan Antara Peubah Asal dengan Komponen Utama

Mencari Skor Komponen Utama

HASIL DAN PEMBAHASAN

Tahap awal penentuan komponen utama adalah mendapatkan akar ciri dan vektor ciri dari matriks ragam peragam atau matriks korelasi. Penelitian ini menggunakan kedua matriks. Berdasarkan data yang diperoleh karena nilai pada data terlalu besar, maka dilakukan transformasi ke bentuk logaritma ($\log (X_i +1)$). Setiap peubah menggunakan satuan pengukuran yang sama sehingga direduksi menggunakan matriks ragam peragam. Data yang telah ditransformasi dapat juga di reduksi dengan menggunakan matriks korelasi.

Adapun pada data peubah yang dapat dianalisis sebanyak 47 peubah, tiga peubah tidak dapat diproses karena selama 2 tahun tidak ada pemasukan. Tiga peubah itu adalah Penerimaan Perusahaan Daerah Air Minum, BANK Pembangunan Daerah, dan Hasil Penjualan Barang Milik Daerah.

Nilai akar ciri menunjukkan besarnya keragaman yang dapat diterangkan oleh masing-masing komponen atau kombinasi linier. Semakin besar nilai akar ciri, maka semakin besar pula persentase keragaman yang diterangkan oleh komponen atau kombinasi linier tersebut, sehingga akar ciri tersusun dari nilai tertinggi sampai nilai yang terendah dimana jumlah seluruh akar ciri sama dengan jumlah seluruh peubah. Data yang direduksi dengan menggunakan matriks ragam peragam menghasilkan nilai akar ciri pertama dan kedua berturut-turut sebesar 41,565 dan 0,022, yang juga

merupakan ragam dari komponen utama itu sendiri. Akar ciri lainnya dapat dilihat pada lampiran 3.a. Vektor ciri yang berkesesuaian dengan akar ciri dapat dilihat pada lampiran 3.b. Sedangkan data yang direduksi dengan menggunakan matriks korelasi menghasilkan nilai akar ciri pertama, kedua, dan ketiga berturut-turut sebesar 33,76; 2,18 dan 1,94. Akar ciri lainnya dapat dilihat pada lampiran 4.a dan vektor ciri yang berkesesuaian dengan akar ciri dapat dilihat pada lampiran 4.b.

Menentukan Banyaknya Komponen Utama

Ada tiga metode untuk menentukan banyaknya komponen utama:

Berdasarkan Kumulatif Proporsi

Dengan ketentuan tersebut, pada matriks ragam peragam diperoleh bahwa dari 47 peubah yang digunakan dapat direduksi menjadi satu komponen utama. Komponen utama pertama tersebut mampu menjelaskan keragaman data asal sebesar 99,801% dari total keragaman dan nilai akar cirinya adalah 41,565. Karena komponen yang dihasilkan hanya satu maka tidak perlu dilakukan rotasi.

Sedangkan pada matriks korelasi diperoleh bahwa dari 47 peubah yang digunakan dapat direduksi menjadi enam komponen utama. Komponen tersebut tidak berkorelasi satu sama lain (orthogonal). Setelah dilakukan proses rotasi, nilai akar ciri mengalami perubahan persentase keragaman data meskipun nilainya tidak jauh berbeda.

Tabel 1. Hasil Rotasi Keragaman Kumulatif Komponen Utama

| Komponen Utama | Akar ciri | Keragaman (%) | Kumulatif Keragaman (%) |
|----------------|-----------|---------------|-------------------------|
| Pertama | 30.38 | 64.63 | 64.63 |
| Kedua | 3.94 | 8.39 | 73.02 |
| Ketiga | 2.90 | 6.16 | 79.18 |
| Keempat | 1.98 | 4.22 | 83.40 |
| Kelima | 1.34 | 2.86 | 86.26 |
| Keenam | 1.19 | 2.53 | 88.78 |

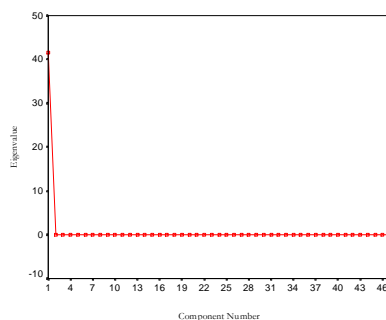
Sumber : Hasil olahan SPSS

Namun keragaman yang dapat dijelaskan oleh keenam komponen utama tersebut sebelum maupun sesudah rotasi tetap sama yaitu sebesar 88,78%.

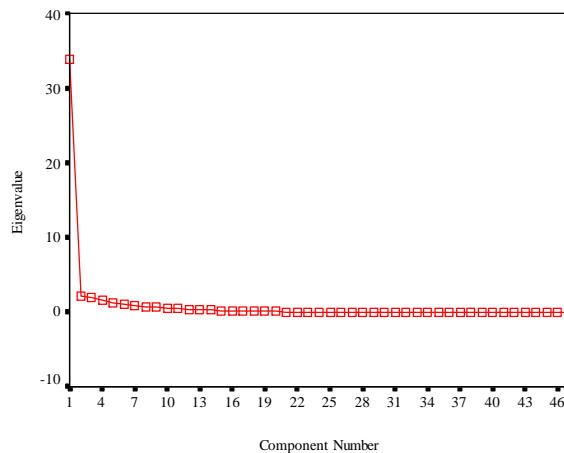
Berdasarkan Akar ciri

Jika data direduksi dengan matriks ragam peragam maka metode ini tidak dapat digunakan. Namun, jika data direduksi menggunakan matriks korelasi maka komponen utama yang d ihasilkan ada enam komponen.

Berdasarkan Scree Plot.



Gambar 1 The Scree Plot Matriks Ragam Peragam



Gambar 2 The Scree Plot Matriks Korelasi

Berdasarkan Gambar 1 plot di atas dapat dilihat bahwa dari komponen pertama ke komponen dua (garis dari sumbu Component Number antara satu dan empat), arah garis menurun dengan cukup tajam. Kemudian dari komponen dua scree mulai merata. Hal ini menunjukkan bahwa satu komponen adalah paling bagus untuk meringkas 47 peubah asal tersebut.

Berdasarkan Gambar 2 plot di atas dapat dilihat bahwa dari komponen utama pertama ke komponen dua (garis dari sumbu Component Number antara satu dan tujuh), arah garis menurun dengan cukup tajam. Kemudian dari komponen dua ke komponen tiga garis masih menurun. Demikian pula dari komponen tiga sampai kekomponen lima, namun dengan slope yang lebih kecil. Slope mulai merata pada komponen keenam, sehingga menunjukkan bahwa lima komponen adalah paling bagus untuk meringkas 47 peubah asal tersebut.

Dari penjelasan di atas maka dapat ditentukan berapa banyak komponen utama yang akan digunakan. Pemilihan berapa banyak komponen yang akan digunakan dapat dilakukan secara subjektif, tergantung dari analisis lanjutan. Dalam kasus penerimaan pendapatan asli daerah kota Bengkulu, jika data direduksi dengan menggunakan matriks ragam peragam komponen utama yang digunakan adalah satu. Hal ini dikarenakan satu komponen utama telah mampu menerangkan total keragaman sebesar 99,801% dari peubah asal. Namun jika data direduksi dengan menggunakan matriks korelasi maka komponen utama yang digunakan adalah empat komponen. Keempat komponen utama tersebut telah mampu menerangkan keragaman data sebesar 83,83% dari peubah asal.

Korelasi Antara Peubah Asal dengan Komponen Utama

Pada analisis komponen utama ini akan diukur korelasi antara peubah asal dengan komponen utama yang terbentuk berdasarkan matriks factor loadings. Komponen utama menghasilkan matriks factor loadings yang tidak berkorelasi satu sama lain dan nilai -nilainya merupakan koefisien korelasi antara peubah asal dengan komponen utama tersebut, dimana ketentuan tanda positif dan negatif diabaikan.

Jika direduksi dengan matriks ragam peragam maka peubah-peubah yang memiliki nilai korelasi tinggi dengan komponen utama pertama adalah Hasil Lelang Barang Milik Daerah (X_{46}), Bagian Laba Perusahaan Milik Daerah (X_{39}), Sumbangan Pihak Ketiga Sektor Pertanian (X_{45}), dan Retribusi Pemanfaatan Alsintan (X_{16}) dengan nilai korelasinya adalah 1. Sedangkan peubah-peubah yang memiliki nilai korelasi rendah dengan komponen utama pertama adalah Retribusi Tanda Daftar Perusahaan (TDP) (X_{31}), Retribusi Surat Izin Usaha Jasa Kontruksi (X_{33}), Retribusi Biro Perjalanan Wisata (X_{37}), Retribusi Leges (X_{26}), dan Retribusi Biaya Cetak KTP dan Akte Kelahiran (X_{12}) yang secara berurutan nilai korelasinya adalah 0.443; 0.257; 0.246; 0.201; 0.056.

Nilai korelasi antara peubah asal dengan komponen utama pertama sangat bervariasi, sehingga dapat dikelompokkan dengan kriteria sebagai berikut:

| | | |
|----------------------|---|-----------------------|
| <i>Sangat Tinggi</i> | <i>jika</i> $K_{i1} \geq 0.996$ | |
| <i>Tinggi</i> | <i>jika</i> $0.950 \leq K_{i1} < 0.995$ | |
| <i>Sedang</i> | <i>jika</i> $0.850 \leq K_{i1} < 0.949$ | $i = 1, 2, \dots, 47$ |
| <i>Rendah</i> | <i>jika</i> $0.650 \leq K_{i1} < 0.849$ | |
| <i>Sangat Rendah</i> | <i>jika</i> $K_{i1} < 0.649$ | |

dimana: K_{i1} = Korelasi peubah ke-i dengan komponen utama pertama

Berdasarkan kriteria diatas maka hasil perhitungan didapat bahwa sembilan peubah masuk kekelompok sangat tinggi, sepuluh peubah masuk kekelompok tinggi, sembilan peubah masuk kekelompok sedang, sepuluh peubah masuk kekelompok rendah dan sembilan peubah masuk kekelompok sangat rendah. Sehingga dapat dikatakan bahwa hasil pengelompokan peubah-peubah tersebut sama rata.

Dari hasil matriks factor loadings dapat dilihat bahwa semua peubah berkorelasi tinggi terhadap komponen utama pertama. Karena komponen yang dihasilkan hanya satu maka tidak (perlu) dirotasi.

Korelasi antara peubah asal dengan komponen utama (sebelum dan sesudah rotasi) bila data direduksi dengan menggunakan matriks korelasi dapat dilihat pada Tabel 5. Meskipun dari proses factor loadings sudah bisa ditetapkan suatu peubah masuk kedalam suatu komponen, namun untuk memperjelas hal itu perlu dilakukan rotasi. Rotasi dilakukan dengan menggunakan rotasi varimax. Nilai keragaman komponen-komponen tersebut tidak berubah sebelum ataupun sesudah rotasi. Dengan rotasi, dapat lebih jelas menafsirkan hubungan peubah-peubah asal dengan masing-masing komponennya. Proses rotasi yang dilakukan sebanyak tujuh kali.

Besarnya ragam dari suatu peubah yang dapat dijelaskan oleh komponen yang terbentuk dapat dilihat dari nilai communalities-nya. Semakin besar nilai communalities-nya maka semakin erat hubungannya dengan komponen yang terbentuk, demikian pula sebaliknya. Pada hasil communalities, ragam dapat dilihat pada extraction dikalikan dengan 100 persen. Bila nilainya besar (> 0.5) maka peubah tersebut erat hubungannya dengan komponen yang terbentuk.

Skor Komponen Utama

Biasanya dalam analisis komponen utama dari p peubah asal dipilih k komponen utama saja yang mampu menerangkan keragaman data cukup tinggi ($k < p$). Untuk tujuan analisis lanjutan, seperti analisis regresi komponen utama atau analisis klaster biasanya dihitung skor komponen utama dari setiap pengamatan. Akan tetapi kalau tujuan analisis komponen utama hanya untuk mereduksi, dari banyak peubah asal menjadi sedikit peubah yang disebut komponen maka perhitungan skor komponen tidak diperlukan. Skor komponen utama umumnya digunakan sebagai bagian dalam eksplorasi data. Pada penelitian ini pengamatan yang digunakan adalah bulan selama dua tahun.

Tabel 2 Korelasi Antara Peubah Asal dengan Komponen Utama

| Peubah | Korelasi Matriks | | | | | | | |
|--------|-------------------------|--------|--------|--------|-------------------------|--------|--------|--------|
| | Komponen Sebelum Rotasi | | | | Komponen Sesudah Rotasi | | | |
| | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| X 1 | -0.545 | -0.383 | -0.113 | 0.370 | -0.505 | -0.470 | 0.160 | -0.056 |
| X 2 | 0.945 | -0.003 | 0.066 | -0.137 | 0.920 | 0.239 | 0.106 | 0.045 |
| X 3 | 0.756 | -0.243 | 0.180 | 0.061 | 0.756 | -0.044 | 0.241 | 0.188 |
| X 4 | 0.984 | -0.010 | -0.017 | 0.024 | 0.930 | 0.193 | 0.253 | -0.058 |
| X 5 | 0.933 | -0.031 | -0.054 | 0.101 | 0.873 | 0.141 | 0.306 | -0.097 |
| X 6 | -0.997 | 0.031 | 0.008 | 0.027 | -0.957 | -0.189 | -0.208 | 0.038 |
| X 7 | 0.989 | -0.005 | -0.013 | -0.052 | 0.948 | 0.215 | 0.183 | -0.044 |
| X 8 | 0.997 | -0.032 | -0.016 | -0.025 | 0.957 | 0.185 | 0.208 | -0.045 |
| X 9 | 0.901 | 0.153 | -0.042 | -0.107 | 0.838 | 0.349 | 0.123 | -0.095 |
| X 10 | 0.689 | -0.105 | -0.546 | 0.056 | 0.691 | -0.065 | 0.113 | -0.540 |
| X 11 | 0.675 | 0.103 | 0.014 | 0.483 | 0.616 | 0.149 | 0.534 | -0.102 |
| X 12 | -0.054 | -0.292 | 0.847 | -0.083 | 0.001 | -0.110 | 0.019 | 0.894 |
| X 13 | 0.966 | -0.041 | 0.207 | -0.079 | 0.931 | 0.223 | 0.186 | 0.180 |
| X 14 | -0.996 | 0.024 | 0.008 | 0.027 | -0.955 | -0.195 | -0.208 | 0.040 |
| X 15 | 0.835 | -0.060 | -0.407 | -0.082 | 0.839 | 0.062 | 0.046 | -0.403 |
| X 16 | -0.997 | 0.024 | 0.008 | 0.031 | -0.957 | -0.195 | -0.204 | 0.039 |
| X 17 | 0.854 | 0.065 | 0.445 | -0.002 | 0.773 | 0.328 | 0.284 | 0.382 |
| X 18 | 0.593 | -0.033 | 0.259 | 0.345 | 0.489 | 0.077 | 0.409 | 0.186 |
| X 19 | 0.911 | 0.084 | 0.147 | -0.049 | 0.845 | 0.312 | 0.205 | 0.094 |
| X 20 | 0.986 | -0.039 | 0.003 | 0.025 | 0.938 | 0.171 | 0.255 | -0.032 |
| X 21 | 0.836 | -0.010 | -0.056 | -0.194 | 0.835 | 0.197 | 0.004 | -0.059 |
| X 22 | 0.957 | -0.036 | -0.030 | 0.020 | 0.912 | 0.162 | 0.238 | -0.063 |
| X 23 | 0.786 | 0.216 | -0.105 | 0.425 | 0.609 | 0.267 | 0.597 | -0.238 |
| X 24 | 0.986 | 0.018 | -0.046 | -0.074 | 0.945 | 0.234 | 0.159 | -0.078 |
| X 25 | 0.908 | 0.178 | -0.003 | 0.075 | 0.800 | 0.346 | 0.307 | -0.088 |
| X 26 | 0.223 | 0.702 | 0.237 | 0.013 | 0.023 | 0.748 | 0.183 | 0.069 |
| X 27 | -0.921 | 0.144 | 0.170 | 0.000 | -0.914 | -0.030 | -0.175 | 0.171 |
| X 28 | -0.815 | 0.318 | -0.013 | -0.028 | -0.845 | 0.127 | -0.188 | -0.043 |
| X 29 | -0.994 | 0.018 | -0.006 | 0.043 | -0.954 | -0.205 | -0.195 | 0.025 |
| X 30 | 0.628 | 0.119 | 0.098 | 0.558 | 0.448 | 0.155 | 0.710 | -0.033 |
| X 31 | 0.453 | 0.644 | 0.073 | -0.263 | 0.318 | 0.766 | -0.059 | -0.049 |
| X 32 | -0.984 | 0.053 | 0.031 | 0.028 | -0.951 | -0.161 | -0.197 | 0.054 |
| X 33 | -0.242 | 0.071 | 0.157 | 0.385 | -0.331 | -0.031 | 0.342 | 0.094 |
| X 34 | 0.873 | -0.124 | 0.030 | 0.089 | 0.838 | 0.059 | 0.284 | 0.008 |
| X 35 | 0.626 | -0.319 | -0.107 | 0.273 | 0.621 | -0.239 | 0.356 | -0.098 |
| X 36 | 0.834 | -0.005 | 0.395 | -0.023 | 0.778 | 0.253 | 0.244 | 0.352 |
| X 37 | 0.279 | 0.741 | -0.196 | 0.193 | 0.049 | 0.680 | 0.299 | -0.385 |
| X 38 | 0.967 | 0.025 | 0.045 | 0.000 | 0.907 | 0.239 | 0.241 | -0.001 |
| X 39 | -0.998 | 0.024 | 0.007 | 0.028 | -0.956 | -0.195 | -0.208 | 0.038 |
| X 41 | 0.919 | -0.054 | -0.053 | -0.135 | 0.912 | 0.163 | 0.076 | -0.058 |
| X 44 | 0.816 | -0.020 | -0.014 | -0.026 | 0.783 | 0.159 | 0.165 | -0.039 |
| X 45 | 0.998 | -0.026 | -0.010 | -0.031 | 0.957 | 0.193 | 0.204 | -0.040 |
| X 46 | 0.998 | -0.026 | -0.009 | -0.032 | 0.957 | 0.195 | 0.203 | -0.039 |
| X 47 | -0.991 | 0.006 | 0.005 | 0.056 | -0.950 | -0.217 | -0.182 | 0.036 |
| X 48 | -0.979 | 0.021 | 0.030 | 0.086 | -0.950 | -0.201 | -0.145 | 0.053 |
| X 49 | -0.997 | 0.015 | 0.010 | 0.045 | -0.957 | -0.206 | -0.192 | 0.040 |
| X 50 | -0.675 | 0.126 | 0.152 | 0.184 | -0.714 | -0.034 | 0.053 | 0.122 |

KESIMPULAN

Hasil yang diperoleh menunjukkan bahwa komponen yang dihasilkan sebanyak 47 komponen sesuai dengan banyaknya peubah asal. Jika data direduksi dengan menggunakan matriks ragam peragam maka komponen utama yang digunakan hanya satu yaitu komponen utama pertama yang menjelaskan keragaman peubah asal sebesar 99,801% dari total keragaman dengan nilai akar ciri 41,565. Komponen utama pertama dapat diidentifikasi sebagai komponen pendapatan asli daerah Kota Bengkulu. Dan jika data direduksi dengan menggunakan matriks korelasi maka komponen utama yang digunakan ada empat komponen. Keempat komponen utama ini mampu menjelaskan keragaman peubah asal sebesar 83,83% dari total keragaman. Penggunaan matriks korelasi memang cukup efektif kecuali pada dua hal. Pertama, secara teori pengujian statistik terhadap akar ciri matriks korelasi jauh lebih rumit dibandingkan penggunaan matriks ragam peragam. Kedua, dengan menggunakan matriks korelasi setiap peubah dipaksakan memiliki ragam yang sama sehingga seringkali tujuan untuk mendapatkan peubah yang kontribusinya paling besar menjadi tidak tercapai.

DAFTAR PUSTAKA

- [1]. Anonim. 1999. Panduan Pengelolaan Data dengan Paket Program Minitab Windows. Edisi Kedua. Jurusan Statistika FMIPA IPB
- [2]. Anonim. 2001. <http://www.gematel.com/edisi40/gema%2utama>.
- [3]. Anonim. 2003. Analisis Peubah Ganda. Jurusan Matematika IPB
- [4]. Anton, H. 1995. Aljabar Linier Elementer. Edisi Kelima. Erlangga. Jakarta
- [5]. Brotodiharjo, R.S. 1998. Pengantar Ilmu Hukum Pajak. Refika Aditama. Bandung
- [6]. Gaspersz, V. 1992. Teknik Analisis Dalam Penelitian Percobaan 2. Tarsito. Bandung.
- [7]. Ichsan, M.H. 1986. Buku Materi Pokok Administrasi Perpajakan. Universitas Terbuka. Depdikbud Jakarta.
- [8]. Johnson, R.A & Wichern, D.W. 2002. Multivariate Analysis Methods and Application. By John Wiley & Sons.
- [9]. James, M.G & Weaver, W. 1987. Aljabar Matriks Untuk Para Insinyur. Edisi Kedua. Erlangga. Jakarta.
- [10]. Mamesah, D.J. 1995. Sistem Administrasi Keuangan Daerah. Gramedia Pustaka Utama. Jakarta.
- [11]. Moehar, D. 2002. Metode Penelitian Sosial Ekonomi. Bumi Aksara. Bandung
- [12]. Santoso, S. 2004. Buku Latihan SPSS Statistik Multivariat. PT Elex Media Komputindo Kelompok Gramedia. Jakarta.
- [13]. Sujanto. 1984. Otonomi Daerah yang Nyata dan Bertanggung Jawab. Cetakan Pertama. Chalia Indonesia. Jakarta.
- [14]. Sugiyono. 2003. Statistika Untuk Penelitian. Alfabeta. Bandung.
- [15]. Supranto, J. 2004. Analisis Multivariat: Arti & Interpretasi. Rineka Cipta. Jakarta
- [16]. Taufiqurrahman. 2003. Skripsi "Analisis Dana Alokasi Umum Propinsi Bengkulu dalam Kerangka Pelaksanaan Otonomi Daerah". UNIB. Bengkulu
- [17]. William, R.D & M.Goldstein. 1984. Multivariate Analysis Methods And Applications. John, W & Sons. Canada.

Perbandingan Hasil Klasifikasi Analisis Diskriminan dan Regresi Logistik Respon Biner Pada Pengklasifikasian Kejadian Bayi dengan Berat Badan Lahir Rendah (BBLR)

Elvy Kurniati¹, Sigit Nugroho², dan Baki Swita²

¹Alumni Jurusan Matematika Fakultas MIPA Universitas Bengkulu

²Staf Pengajar Jurusan Matematika Fakultas MIPA Universitas Bengkulu

ABSTRAK

Analisis diskriminan adalah analisis multivariat yang bertujuan untuk mengklasifikasikan suatu pengamatan baru ke dalam salah satu dari kelompok (grup) yang telah diketahui. Dalam penerapannya memerlukan asumsi multivariat normal dan kesamaan struktur matriks varian kovarian antar grup, tetapi kedua asumsi ini tidak selalu dapat dipenuhi. Regresi logistik respon biner adalah bentuk khusus regresi logistik dengan respon biner dimana persamaannya menghasilkan nilai peluang yang dipakai sebagai dasar untuk klasifikasi. Analisis ini tidak memerlukan asumsi seperti halnya analisis diskriminan sehingga dapat diterapkan dalam berbagai skala data. Tujuan penelitian ini adalah membentuk fungsi klasifikasi analisis diskriminan dan regresi logistik respon biner serta membandingkan hasil klasifikasi tersebut dalam kasus kejadian bayi dengan berat badan lahir rendah (BBLR). Hasil yang diperoleh menunjukkan bahwa asumsi multivariat tidak dapat dipenuhi dan matriks varian kovarian dipenuhi. Hasil penelitian ini menunjukkan juga apabila variabel penjelas berskala kategori, regresi logistik respon biner menghasilkan ketepatan klasifikasi yang tinggi dibandingkan dengan analisis diskriminan.

Kata kunci : Analisis Diskriminan, Regresi Logistik Respon Biner, Klasifikasi

PENDAHULUAN

Analisis diskriminan adalah analisis multivariat yang diterapkan untuk memodelkan hubungan antara satu variabel respon yang bersifat dikotomis atau multikotomis maupun kualitatif dengan variabel penjelas yang bersifat kuantitatif (Hair *et al.*, 1998). Analisis diskriminan ini dilakukan berdasarkan perhitungan statistik terhadap kelompok (grup) yang terlebih dahulu diketahui secara jelas. Apabila terdapat dua atau lebih populasi yang telah diukur dalam beberapa karakter (x_1, x_2, \dots, x_k) , maka dapat dibangun fungsi linear tertentu dari pengukuran itu, dimana fungsi tersebut merupakan fungsi pembeda (pemisah) terbaik bagi populasi-populasi yang diamati. Fungsi linier yang dibangun itu disebut sebagai fungsi diskriminan (*discriminant function*) (Gaspersz, 1992).

Tujuan dari analisis diskriminan ini adalah untuk menghasilkan suatu aturan yang menggambarkan pemisahan beberapa kelompok (grup) pengamatan, serta mengklasifikasikan suatu pengamatan baru ke dalam salah satu dari kelompok (grup) yang telah diketahui. Oleh karena itu, analisis diskriminan ini dapat dipergunakan sebagai metode pengklasifikasian. Dalam penerapannya, fungsi diskriminan ini mengasumsikan bahwa data pada masing-masing kategori/kelompok berdistribusi normal multivariat dan mempunyai struktur matriks varian kovarian yang sama (Hair *et al.*, 1998).

Regresi logistik respon biner adalah bentuk khusus analisis regresi dengan respon biner dan variabel penjelas yang dapat terdiri dari data kontinu, kategori atau campuran antara keduanya. Persamaan regresi logistik respon biner ini tidak menghasilkan nilai pada variabel respon, namun menghasilkan nilai peluang kejadian pada variabel respon. Nilai peluang ini yang dipakai sebagai ukuran untuk mengklasifikasikan pengamatan. Regresi logistik respon biner tidak

memerlukan asumsi distribusi multivariat normal dan kesamaan varian kovarian dalam masing-masing grup (Hosmer & Lemeshow dalam Wibowo, 2002). Sehingga metode ini cukup tahan (*robust*) untuk dapat diterapkan dalam berbagai skala/keadaan data (Hair *et al.*, 1998).

Berdasarkan penjelasan diatas, dapat diketahui bahwa salah satu kegunaan kedua metode tersebut adalah pengklasifikasian, walaupun kedua metode ini berbeda dalam asumsi atau metode estimasi serta inferensi parameternya. Sehingga dengan melihat hasil perbandingan klasifikasi antara analisis diskriminan dan regresi logistik respon biner dapat diketahui pada situasi bagaimana kedua metode tersebut sebaiknya digunakan. Salah satu contoh penerapannya adalah untuk melihat ketepatan klasifikasi kejadian bayi dengan berat badan lahir rendah (BBLR) terhadap faktor-faktor yang mempengaruhinya dengan menggunakan analisis diskriminan dan regresi logistik respon biner.

Berdasarkan uraian diatas maka penulis tertarik untuk mengetahui ketepatan pengklasifikasian kejadian bayi dengan BBLR. Oleh karena variabel respon dalam penelitian ini adalah kejadian bayi dengan BBLR yang berupa data kategori (dua kategori atau biner) dan variabel penjelas adalah umur ibu, parietas, jarak kelahiran, *Antenatal Care* (ANC) serta kenaikan berat badan merupakan data kuantitatif maka analisis ini dapat digunakan analisis diskriminan dan regresi logistik respon biner dalam menganalisis ketepatan pengklasifikasian kejadian bayi dengan BBLR.

TINJAUAN PUSTAKA

a. Faktor-faktor yang Mempengaruhi BBLR

Bayi berat badan lahir rendah (BBLR) adalah bayi baru lahir yang berat badan lahirnya pada saat kelahiran kurang dari 2500 gram. Faktor-faktor yang mempengaruhi bayi BBLR ini antara lain (Syafrie dkk., 2004) umur ibu, parietas, jarak kelahiran dengan bayi terdahulu, *antenatal care* (ANC) yang merupakan pelayanan kesehatan bagi ibu hamil dan janinnya oleh tenaga profesional meliputi pemeriksaan kehamilan sesuai dengan standar pelayanan yaitu minimal 4 kali pemeriksaan selama kehamilan, 1 kali pada trimester I, 1 kali pada trimester II dan 2 kali pada trimester III, serta kenaikan berat badan ibu selama hamil.

b. Analisis Diskriminan

Analisis diskriminan adalah analisis multivariat yang diterapkan untuk memodelkan hubungan antara satu variabel respon yang bersifat dikotomis atau multikotomis maupun kualitatif dengan variabel penjelas yang bersifat kuantitatif (Hair *et al.*, 1998). Tujuan dari analisis diskriminan ini adalah untuk menghasilkan suatu aturan yang menggambarkan pemisahan beberapa kelompok (grup) pengamatan, serta mengklasifikasikan suatu pengamatan baru ke dalam salah satu dari kelompok (grup) yang telah diketahui.

Jika \mathbf{x} adalah pengamatan baru yang tidak diketahui asalnya, maka dapat dihitung skor diskriminan linier yaitu:

$$W_{ij} = \mathbf{x}'\mathbf{S}_{pl}^{-1}(\bar{\mathbf{x}}_i - \bar{\mathbf{x}}_j) - \frac{1}{2}(\bar{\mathbf{x}}_i + \bar{\mathbf{x}}_j)'\mathbf{S}_{pl}^{-1}(\bar{\mathbf{x}}_i - \bar{\mathbf{x}}_j) \quad (1)$$

Sehingga kriteria klasifikasi berdasarkan statistik W (*Wald-Anderson*) adalah:

$$\text{golongan } \mathbf{x} \text{ ke grup } i \text{ jika } W_{ij} > 0 \text{ untuk semua } j \neq i \quad (2)$$

dengan catatan $W_{ij} = -W_{ji}$ (Morrison, 1978).

Keterangan:

$\bar{\mathbf{x}}_i$ = vektor rata-rata sampel grup ke- i

$\bar{\mathbf{x}}_j$ = vektor rata-rata sampel grup ke- j

W_{ij} = Fungsi diskriminan yang akan menggolongkan individu ke grup i atau grup j

Menurut Gazpert (1992), jika hanya ada dua grup maka kriteria klasifikasi berdasarkan satatistik W adalah:

1. Golongkan \mathbf{x} ke dalam grup 1 jika $W_{12} > 0$
2. Golongkan \mathbf{x} ke dalam grup 2 jika $W_{12} \leq 0$

Salah satu cara untuk menghindari bias adalah membagi sampel menjadi dua bagian yaitu *training sample* yang digunakan untuk membentuk aturan klasifikasi dan *validation sample* yang digunakan untuk mengevaluasi fungsi klasifikasi (Rhencher, 1995). Hair *et al.* (1998) menyatakan, tidak ada acuan pasti dalam membagi sampel menjadi *training sample* dan *validation sample*. Proporsi pembagian sampel ini tidak harus sama besar untuk masing-masing bagian misalnya 60% dan 40%, 75% dan 25% atau yang lainnya. Wibowo (20 02) membuat proporsi pembagian sampel sebesar 80% dan 20% dan Suhaeni (2005) sebesar 60% dan 40%. Proses validasi terhadap fungsi diskriminan yang terbentuk di *training sample* harus dilakukan berkali-kali, yang jelas tidak cukup hanya sekali (Supranto, 200 4).

c. Analisis Regresi Logistik Respon Biner

Analisis regresi logistik respon biner adalah bentuk khusus analisis regresi dengan respon biner dan variabel penjelas yang dapat terdiri dari data kontinu, kategori atau campuran antara keduanya (Hosmer & Lemeshow dalam Wibowo, 2002). Regresi logistik dengan respon biner ini merupakan regresi logistik dimana hanya terdapat dua kemungkinan dari nilai Y . Analisis ini digunakan untuk melihat pengaruh sejumlah variabel penjelas x_1, x_2, \dots, x_k terhadap variabel respon Y yang berupa variabel respon biner atau juga untuk memprediksi nilai suatu variabel respon Y (yang berupa variabel biner) berdasarkan nilai variabel-variabel penjelas x_1, x_2, \dots, x_k .

Model regresi logistik didefinisikan sebagai berikut, misalkan Y_1, \dots, Y_n adalah variabel respon Bernoulli dan π_i sebagai nilai harapan dari Y_i yaitu $\pi_i = E[Y_i] = P[Y_i = 1]$. Nilai harapan π_i dapat dinyatakan dalam bentuk dari variabel penjelas $x_{i,1}, x_{i,2}, \dots, x_{i,k}$ yaitu

$$\pi_i = \frac{1}{1 + \exp\left(-\beta_0 - \sum_{j=1}^k \beta_j x_{i,j}\right)} \quad (3)$$

dimana $\beta = (\beta_1, \dots, \beta_k)$ adalah parameter yang tidak diketahui. Jika persamaan (3) dilakukan transformasi logit, akan didapatkan hubungan linier logit (π_i).

Sehingga didapat persamaan logit:

$$\text{logit}(\pi_i) = \ln\left(\frac{\pi_i}{1 - \pi_i}\right) = \beta_0 + \sum_{j=1}^k \beta_j x_{i,j} \quad (4)$$

Persamaan (4) dinamakan model logit atau logistik dan merupakan fungsi yang linier dalam parameternya.

Penduga parameter dalam regresi logistik respon biner menggunakan maximum likelihood (Larsen, 2006). Pada dasarnya metode maksimum likelihood memberikan nilai dugaan parameter dengan memaksimumkan suatu fungsi likelihood bagi β dengan iterasi Newton Raphson. Estimasi maksimum likelihood merupakan pendekatan dari estimasi *Weighted Least Square*, dimana matrik pembobotnya berubah setiap putaran. Proses menghitung estimasi maksimum likelihood ini disebut juga sebagai *Iteratif Reweighted Least Square*.

Nilai *odds ratio* merupakan besaran yang digunakan untuk melihat perbandingan masing-masing kategori dari variabel penjelas dalam menerangkan variabel respon. *Odds ratio* biasanya dilambangkan dengan θ didefinisikan sebagai ratio dari odds untuk $x = 1$ terhadap $x = 0$ atau menunjukkan besarnya perbedaan nilai variabel respon ketika variabel penjelas $(x + 1)$ dan nilai variabel respon ketika variabel penjelas x , untuk setiap nilai x . Ratio antara dua odds didefinisikan sebagai berikut:

$$\theta = \frac{\pi(1)/1 - \pi(1)}{\pi(0)/1 - \pi(0)} \quad (4)$$

Pemilihan antara kedua metode tersebut biasanya tergantung pada asumsi yang diperlukan oleh kedua metode tersebut. Analisis diskriminan mengasumsikan data berdistribusi multivariat normal, sedangkan regresi logistik respon biner tidak mengasumsikan data harus berdistribusi tertentu.

Penelitian dengan dua metode ini telah dilakukan oleh Wibowo (2002), dimana apabila variabel penjelas bersifat campuran antara kontinu dan kategori dalam studi kasus data kerusakan *wall tile*, analisis diskriminan sebaiknya digunakan karena akan menghasilkan ketepatan yang lebih tinggi dibanding regresi logistik respon biner apabila tidak ada pelanggaran asumsi. Apabila variabel penjelas bersifat kategori dalam studi kasus data faktor resiko dan kualitas penderita penyakit stroke, regresi logistik respon biner sebaiknya dipergunakan karena akan menghasilkan ketepatan klasifikasi yang tinggi dibanding analisis diskriminan dan apabila variabel prediktor bersifat kontinu dalam studi kasus data demografi, analisis diskriminan sebaiknya dipergunakan karena akan menghasilkan ketepatan klasifikasi yang lebih tinggi dibanding regresi logistik respon biner meskipun ada pelanggaran asumsi, dengan catatan tidak banyak data yang outlier.

METODE PENELITIAN

Jenis penelitian ini merupakan penelitian terapan (*applied research*), yaitu penelitian yang dimaksudkan untuk menerapkan suatu teori dalam kaitannya dengan pemanfaatannya dalam bidang kesehatan. Sampel yang digunakan dalam penelitian ini adalah *purposif* sampel yaitu sebanyak 123 responden ibu yang melahirkan bayi dengan BBLR dan tidak BBLR. Data yang digunakan adalah data sekunder, yang diperoleh dari dokumentasi dari klinik bersalin Tiara Sella kota Bengkulu. Gambaran variabel-variabel yang akan diteliti sebagai berikut:

Tabel 3.1 Gambaran Variabel-variabel yang Akan Diteliti

| Variabel | Label | Skala | Kategori |
|----------------------------|----------------------|---------|-----------------|
| Respon (Y) | Kejadian BBLR | Nominal | 1= ya |
| | | | 0= tidak |
| Penjelas (X ₁) | Umur ibu | Ordinal | 1= < 20 dan >35 |
| | | | 2= 20-35* |
| (X ₂) | Parietas | Ordinal | 1= <2 dan >3 |
| | | | 2= 2-3* |
| (X ₃) | Jarak Kelahiran | Ordinal | 1= <24 dan >36 |
| | | | 2= 24-36* |
| (X ₄) | Kenaikan Berat Badan | Ordinal | 1= < 10 |
| | | | 2= >=10* |
| (X ₅) | Frekuensi ANC | Ordinal | 1= < 4 kali |
| | | | 2= ≥ 4 kali* |

Keterangan: * merupakan kategori pembanding

Tahap-tahap analisis data dalam penelitian ini sebagai berikut:

1. Menguji signifikansi secara parsial terhadap variabel-variabel yang ikut dalam analisis secara parsial.
2. Menguji apakah asumsi multivariat normal dan kesamaan matriks varian kovarian terpenuhi oleh data pada masing-masing grup respon.
3. Membagi data menjadi 2 bagian yaitu *training sample* dan *validation sample* dengan proporsi 75% dan 25%, 80% dan 20%, 60 dan 40%. Data yang masuk ke *training sample* maupun *validation sample* dipilih secara acak.
4. Membentuk model klasifikasi dengan analisis diskriminan dan regresi logistik menggunakan data pada *training sample*. Model klasifikasi yang telah diperoleh ini selanjutnya divalidasi dengan data pada *validation sample*. Jika ketepatan klasifikasi dari kedua sampel hampir sama besar, dikatakan fungsi diskriminan dan logistik respon biner dari *training sample* sudah valid.

HASIL DAN PEMBAHASAN

Hasil uji signifikansi secara parsial, dari lima variabel ternyata ada empat variabel yang memberikan pengaruh secara signifikan (Sig. < 0,05) terhadap kejadian bayi dengan BBLR yaitu umur ibu, jarak kelahiran, kenaikan berat badan serta frekuensi ANC. Variabel parietas ternyata tidak berbeda secara signifikan (Sig. > 0,05).

Untuk selanjutnya, data yang melibatkan semua variabel dalam analisis termasuk variabel parietas disebut sebagai data 1 dan data yang melibatkan semua variabel yang berpengaruh secara signifikan (tanpa variabel parietas) disebut sebagai data 2.

Analisis Data 1

Berdasarkan hasil pengujian asumsi untuk data 1 ini, asumsi multivariat normal tidak dapat dipenuhi sedangkan matriks varian kovarian dapat dipenuhi. Penelitian ini akan dilihat hasil klasifikasi analisis diskriminan dan regresi logistik respon biner jika variabel berskala kategori dan asumsi tidak mutlak dipenuhi dalam kasus kejadian bayi dengan BBLR.

Hasil penerapan analisis diskriminan dan regresi logistik respon biner dari data 1, dengan proporsi 80% dan 20% untuk *training* dan *validation sample* menghasilkan fungsi diskriminan dan regresi logistik respon biner seperti tertera pada tabel 4.1. Dari tabel tersebut dapat dilihat, pada data *training sample* regresi logistik respon biner memiliki ketepatan klasifikasi yang tinggi (85,7%) dibandingkan analisis diskriminan (83,7%). Sedangkan pada *validation sample* regresi logistik juga menghasilkan ketepatan klasifikasi yang tinggi (80%) dibanding analisis diskriminan (72%). Penilaian valid atau tidaknya fungsi diskriminan maupun regresi logistik respon biner adalah jika ketepatan dari *training* dan *validation sample* sudah sama besar (Santoso, 2004), artinya selisih dari keduanya seminimal mungkin. Jika dilihat dari hasil kedua metode tersebut, selisih yang paling minimum adalah metode analisis regresi logistik respon biner. Jadi, metode analisis regresi logistik respon biner dapat memberikan hasil klasifikasi yang lebih baik daripada analisis diskriminan.

Tabel 4.1 Koefisien Fungsi Diskriminan dan Regresi Logistik Respon Biner 1a

| Variabel | Analisis Diskriminan | Regresi Logistik Respon Biner |
|-----------------------|----------------------|-------------------------------|
| Umur Ibu | 0,9924 | 1,8987 |
| Parietas | 0,5077 | 0,9570 |
| Jarak Kelahiran | 0,5919 | 1,8241 |
| Kenaikkan Berat Badan | 1,9893 | 2,9315 |
| frekuensi ANC | 1,3926 | 2,8564 |
| Konstanta | -8,6852 | -7,0504 |

Tabel 4.2 Hasil Klasifikasi Analisis Diskriminan dan Regresi Logistik Respon Biner 1a

| | | | Analisis Diskriminan Predicted Group | | Ketepatan Klasifikasi |
|------------|--------|------------|--|------|--------------------------|
| | | | Tidak BBLR | BBLR | |
| Training | Actual | Tidak BBLR | 65 | 12 | 83,70% |
| Sample | Group | BBLR | 4 | 17 | |
| | | | Tidak BBLR | BBLR | |
| Validation | Group | Tidak BBLR | 15 | 5 | 72% |
| Sample | Actual | BBLR | 2 | 3 | |
| | | | Regresi Logistik Respon Biner Predicted Group | | Ketepatan Klasifikasi |
| | | | Tidak BBLR | BBLR | |
| Training | Actual | Tidak BBLR | 71 | 6 | 85,70% |
| Sample | Group | BBLR | 8 | 13 | |
| | | | Tidak BBLR | BBLR | |
| Validation | Group | Tidak BBLR | 19 | 1 | 80% |
| Sample | Actual | BBLR | 4 | 1 | |

Kemudian, hasil penerapan analisis diskriminan dan regresi logistik respon biner dari data 1, dengan proporsi 75% dan 25% akan menghasilkan fungsi diskriminan dengan ketepatan klasifikasi sebesar 84,8% dan regresi logistik respon biner dengan ketepatan klasifikasi 87% pada *training sample* seperti tertera pada tabel 4.4. Jika dilihat dari hasil kedua metode tersebut, metode analisis regresi logistik respon biner masih memberikan hasil klasifikasi yang lebih tinggi daripada analisis diskriminan.

Tabel 4.3 Koefisien Fungsi Diskriminan dan regresi Logistik Respon Biner 1b

| Variabel | Analisis | |
|-----------------------|-------------|-------------------------------|
| | Diskriminan | Regresi Logistik Respon Biner |
| Umur Ibu | 1,013 | 1,993 |
| Parietas | 0,424 | 0,771 |
| Jarak Kelahiran | 0,531 | 1,737 |
| Kenaikkan Berat Badan | 2,099 | 3,172 |
| frekuensi ANC | 1,333 | 2,882 |
| Konstanta | -8,640 | -7,049 |

Tabel 4.4 Hasil Klasifikasi Analisis Diskriminan dan Regresi Logistik Respon Biner 1b

| | | | Analisis Diskriminan Predicted Group | | Ketepatan Klasifikasi |
|------------|--------|------------|---|------|--------------------------|
| | | | Tidak BBLR | BBLR | |
| Training | Actual | Tidak BBLR | 62 | 11 | 84,8% |
| Sample | Group | BBLR | 3 | 16 | |
| | | | Tidak BBLR | BBLR | |
| Validation | Group | Tidak BBLR | 17 | 7 | 71% |
| Sample | Actual | BBLR | 2 | 5 | |

| | | | Regresi Logistik Respon Biner Predicted Group | | Ketepatan Klasifikasi |
|----------------------|--------|------------|--|------|--------------------------|
| | | | Tidak BBLR | BBLR | |
| Training Sample | Actual | Tidak BBLR | 68 | 5 | 87% |
| | Group | BBLR | 7 | 12 | |
| | | | Tidak BBLR | BBLR | |
| Validation Sample | Group | Tidak BBLR | 22 | 2 | 77,4% |
| | Actual | BBLR | 5 | 2 | |

Jika data 1 dengan proporsi 60% dan 40%, maka menghasilkan fungsi diskriminan dan regresi logistik respon biner seperti tertera pada tabel 4.5. Dari tabel 4.9 dapat dilihat, pada data *training sample* regresi logistik respon biner memiliki ketepatan klasifikasi yang tinggi (86,5%) dibandingkan analisis diskriminan (83,8%). Sedangkan pada *validation sample* regresi logistik respon biner juga menghasilkan ketepatan klasifikasi yang tinggi (79,6%) dibanding analisis diskriminan (71,4%).

Tabel 4.5 Koefisien Fungsi Diskriminan dan Regresi Logistik Respon Biner 1c

| Variabel | Analisis | |
|-----------------------|-------------|-------------------------------|
| | Diskriminan | Regresi Logistik Respon Biner |
| Umur Ibu | 1,113 | 2,114 |
| Parietas | 0,321 | 0,382 |
| Jarak Kelahiran | 0,528 | 1,558 |
| Kenaikkan Berat Badan | 2,166 | 3,153 |
| frekuensi ANC | 1,116 | 2,395 |
| Konstanta | -8,424 | -6,318 |

Tabel 4.6 Hasil Klasifikasi Analisis Diskriminan dan Regresi Logistik Respon Biner 1c

| | | | Analisis Diskriminan Predicted Group | | Ketepatan Klasifikasi |
|----------------------|--------|------------|--|------|--------------------------|
| | | | Tidak BBLR | BBLR | |
| Training Sample | Actual | Tidak BBLR | 49 | 9 | 83,8% |
| | Group | BBLR | 3 | 13 | |
| | | | Tidak BBLR | BBLR | |
| Validation Sample | Group | Tidak BBLR | 27 | 12 | 71,4% |
| | Actual | BBLR | 2 | 8 | |
| | | | Regresi Logistik Respon Biner Predicted Group | | Ketepatan Klasifikasi |
| | | | Tidak BBLR | BBLR | |
| Training Sample | Actual | Tidak BBLR | 53 | 5 | 86,6% |
| | Group | BBLR | 5 | 11 | |
| | | | Tidak BBLR | BBLR | |
| Validation Sample | Group | Tidak BBLR | 36 | 3 | 79,6% |
| | Actual | BBLR | 7 | 3 | |

Jika dilihat dari hasil kedua metode tersebut, metode analisis regresi logistik respon biner dapat memberikan tingkat klasifikasi yang tinggi daripada analisis diskriminan. Ini berarti bahwa data kategori dengan metode analisis regresi logistik respon biner dapat memberikan hasil klasifikasi yang lebih baik daripada analisis diskriminan.

Analisis Data 2

Berdasarkan hasil pengujian asumsi untuk data 2 ini, asumsi multivariat normal tidak dapat dipenuhi sedangkan matriks varian kovarian dapat dipenuhi. Penelitian ini akan dilihat hasil klasifikasi analisis diskriminan dan regresi logistik respon biner jika variabel berskala kategori dan asumsi tidak mutlak dipenuhi dalam kasus kejadian bayi dengan BBLR

Hasil penerapan analisis diskriminan dan regresi logistik respon biner dari data 2, dengan proporsi 80% dan 20% untuk *training* dan *validation sample* akan menghasilkan fungsi diskriminan dan regresi logistik respon biner seperti tertera pada tabel 4.7. Dari tabel tersebut dapat dilihat, pada data *training sample* regresi logistik respon biner memiliki ketepatan klasifikasi yang tinggi (85,7%) dibandingkan analisis diskriminan (81,6%). Sedangkan pada *validation sample* regresi logistik juga menghasilkan ketepatan klasifikasi yang tinggi (80%) dibanding analisis diskriminan (64%). Regresi logistik respon biner juga memberikan selisih yang paling minimum antara *training* dan *validation sample*. Jadi, metode analisis regresi logistik respon biner dapat memberikan hasil klasifikasi yang lebih baik daripada analisis diskriminan.

Tabel 4.7 Hasil Klasifikasi Analisis Diskriminan dan Regresi Logistik Respon Biner

| Analisis Diskriminan | | | |
|--|-------------|-------------|-------------|
| Ketepatan Klasifikasi | | | |
| | 80%-20%(2a) | 75%-25%(2b) | 60%-40%(2c) |
| Training Sample | 83,80% | 82,60% | 81,60% |
| Validation Sample | 75,50% | 74,20% | 64% |
| Analisis Regresi Logistik Respon Biner | | | |
| Ketepatan Klasifikasi | | | |
| | 80%-20%(2a) | 75%-25%(2b) | 60%-40%(2c) |
| Training Sample | 86,5% | 87% | 85,7% |
| Validation Sample | 81,6% | 80,6% | 80% |

Kemudian, hasil penerapan analisis diskriminan dan regresi logistik respon biner dari data 2, dengan proporsi 75% dan 25% akan menghasilkan fungsi diskriminan dengan ketepatan klasifikasi sebesar 82,6% dan regresi logistik respon biner dengan ketepatan klasifikasi 87%. Pada *validation*

sample analisis diskriminan menghasilkan ketepatan klasifikasi 74,2% sedangkan regresi logistik respon biner 80,6%. Jika dilihat dari hasil kedua metode tersebut, metode analisis regresi logistik respon biner masih memberikan hasil klasifikasi yang lebih baik daripada analisis diskriminan.

Jika data 2 dengan proporsi 60% dan 40%, pada *training sample* regresi logistik respon biner memiliki ketepatan klasifikasi yang tinggi (86,5%) dibandingkan analisis diskriminan (83,8%). Sedangkan pada *validation sample* regresi logistik juga menghasilkan ketepatan klasifikasi yang tinggi (81,6%) dibanding analisis diskriminan (75,5%). Jika dilihat dari hasil kedua metode tersebut, metode analisis regresi logistik respon biner dapat memberikan tingkat klasifikasi yang tinggi daripada analisis diskriminan.

Dari berbagai cara analisis seperti tersebut diatas, dapat dikatakan bahwa data 1 dan 2 yang mana variabelnya berskala kategori, metode analisis regresi logistik respon biner akan memberikan hasil ketepatan klasifikasi yang tinggi bila dibandingkan dengan diskriminan. Hal ini bisa terjadi dimungkinkan karena semua variabel penjelasnya berskala kategori sehingga penyebaran datanya sangat menyimpang dari bentuk multivariat normal. Menurut Wibowo (2002) pelanggaran asumsi multivariat normal pada analisis diskriminan biasanya menghasilkan tingkat ketepatan klasifikasi yang rendah.

Dari analisis data 1, ternyata ketepatan klasifikasi data 1b (84,8%-87%) lebih tinggi bila dibanding dengan data 1a (83,7%-85,7%) dan 1c (83,8%-86,5%), namun data 1a memberikan hasil klasifikasi yang lebih baik karena selisih antara ketepatan klasifikasi data pada *training* dan *validation sample* lebih minimum dibanding data 1b dan 1c. Sedangkan dari analisis data 2, ternyata ketepatan klasifikasi data 2b (82,6%-87%) juga lebih tinggi bila dibanding dengan data 2a (81,6%-85,7%) dan 2c (83,8%-86,5%). Hal ini berarti bahwa ternyata cara pembagian sampel dapat memberikan pengaruh terhadap hasil analisis diskriminan dan regresi logistik respon biner. Hasil analisis ini menunjukkan regresi logistik respon biner yang menghasilkan ketepatan klasifikasi yang tinggi dengan proporsi 75%-25%.

Dari berbagai cara analisis yang telah dilakukan, maka yang memberikan ketepatan klasifikasi yang tinggi adalah analisis regresi respon biner dengan data 2 (tanpa variabel parietas) yang mempunyai proporsi 75%-25%, yaitu sebesar 87% dari *training sample* dan 80,6% dari *validation sample*.

Hasil penelitian ini menunjukkan bahwa kejadian bayi dengan BBLR dipengaruhi oleh umur ibu, jarak kelahiran, kenaikan berat badan serta frekuensi ANC. Untuk melihat seberapa besar kecenderungan kejadian BBLR berdasarkan variabel-variabel yang mempengaruhinya digunakan *odds ratio* (θ).

Tabel 4.8 Output *Odds Ratio* dalam Analisis Regresi Logistik Respon Biner Data 2b

| Variabel | Exp(B) |
|----------------------|--------|
| Umur | 7,036 |
| Jarak Kelahiran | 6,118 |
| Kenaikan Berat Badan | 22,268 |
| Frekuensi ANC | 18,341 |
| Konstan | 0,002 |

Nilai rasio kecenderungan masing-masing variabel yang berpengaruh terhadap kejadian bayi dengan BBLR yang tertera pada tabel 4.13 adalah umur ibu sebesar 7,036 berarti bahwa peluang bagi responden yang berusia kurang dari 20 tahun dan besar dari 35 tahun berisiko melahirkan bayi dengan BBLR 7,036 kali lebih tinggi dibandingkan responden yang berumur 20 -35 tahun. Nilai rasio untuk jarak kelahiran sebesar 6,118 yang mana memperlihatkan bahwa peluang bagi responden yang jarak kelahiran kurang dari 24 bulan dan besar dari 36 bulan berpeluang melahirkan BBLR 6,118 kali lebih tinggi dibandingkan responden dengan jarak kelahiran 24 -36 bulan. Nilai rasio untuk kenaikan berat

badan sebesar 22,268 artinya responden dengan kenaikan berat badan kurang dari 10 kg memiliki kecenderungan memperoleh BBLR sebesar 22,268 kali dibandingkan responden yang kenaikan berat badannya lebih dari 10 kg. Nilai rasio untuk frekuensi ANC sebesar 18,341 bermakna bahwa responden dengan frekuensi ANC kurang dari 4 kali memiliki kecenderungan memperoleh BBLR sebesar 18,341 kali dibandingkan responden dengan frekuensi ANC lebih dari 4 kali.

Kesimpulan

Berdasarkan analisis dan pembahasan pada bab sebelumnya, maka dapat disimpulkan yaitu apabila variabel penjelas berskala kategori dalam kasus kejadian bayi dengan BBLR ini, regresi logistik respon biner menghasilkan ketepatan klasifikasi yang tinggi dibandingkan dengan analisis diskriminan. Variabel-variabel yang mempengaruhi kejadian bayi dengan BBLR serta nilai rasionya adalah umur ibu dengan nilai ratio sebesar 7,036, jarak kelahiran (6,118), kenaikan berat badan (22,268) serta frekuensi ANC (18,341).

DAFTAR PUSTAKA

- [1]. Anonim. 2002. *Logistic Regression*. <http://online.sfsu.edu/~efc/classes/biol710/logistic/logisticreg.htm>
- [2]. Anonim. 2004. *Estimation and Hypothesis Testing for Logistic Regression*. <http://courses.washington.edu/b515/113.pdf>
- [3]. Anonim. 2005. *Discriminant Function Analysis*. <http://www.statsoft.com/textbook/stdiscan.html>
- [4]. Anonim. 2005. *Indikator Kesejahteraan Rakyat Propinsi Bengkulu*. Badan Pusat Statistik. Bengkulu.
- [5]. Fox, J. 2005. *Maximum Likelihood Estimation of the Logistic Regression Model*. www.socserv.mcmaster.ca/jfox/Courses/UCLA/logistic-regression-notes.pdf
- [6]. Garson, G.D. 2006. *Logistic Regression*. <http://www2.chass.ncsu.edu/garson/pa765/logistic.htm>
- [7]. Gaspersz, V. 1992. *Teknik Analisis Dalam Penelitian Percobaan*. Tarsito. Bandung.
- [8]. Hair, J.F. et al. 1998. *Multivariate Data Analysis*. Prentice Hall. New Jersey.
- [9]. Larsen, P.V. 2006. *Logistic Regression*. <http://statmaster.sdu.dk/courses/st111/module14/module.pdf>
- [10]. Morrison, D.F. 1978. *Multivariate Statistical Methods*. International Student Edition, McGrawhill.
- [11]. Permata, A. 2006. *Analisis Regresi Logistik Status Kemiskinan Penduduk Kota Bengkulu Tahun 2004*. Skripsi FMIPA Universitas Bengkulu.
- [12]. Rhencher, A.C. 1995. *Method of Multivariate Analysis*. John Wiley and Sons. New York.
- [13]. Ridgeway, G. 2004. *Maximum Likelihood and Logistic Regression*. www.i-pensieri.com/gregr/ModernPrediction/L2logistic.pdf
- [14]. Santoso, S. 2002. *SPSS Statistik Multivariat*. Gramedia. Jakarta.
- [15]. Seber, G. A. F. 1984. *Multivariate Observation*. John Wiley and Sons. New York.
- [16]. Sitohang, N.A. 2004. *Asuhan Keperawatan Pada Bayi Berat Badan Lahir Rendah*. <http://library.usu.ac.id/download/fk/04006076.pdf>
- [17]. Suhaeni, C. 2005. *Analisis Diskriminan Untuk Prediksi Indeks Prestasi Mahasiswa Pada Semester Pertama*. Skripsi FMIPA Universitas Bengkulu.
- [18]. Supranto, J. 2004. *Analisis Multivariat: Arti dan Interpretasi*. Rineka Cipta. Jakarta.
- [19]. Syafrie, dkk.. 2004. *Hubungan Pelayanan Antenatal Terhadap Kejadian Bayi Berat badan Lahir Rendah di Kabupaten Purworejo*. www.akatiga.or.id/black_paper/van_bem_grijns.pdf

- [20]. Wibowo, W. 2002. *Perbandingan Hasil Klasifikasi Analisis Diskriminan dan Regresi Logistik Pada Pengklasifikasian Data Respon Biner*. Kappa, Vol 3, No 1:36-45. [www.fmipa.its.ac.id/isi%20mipajurnal/jurnal/KAPPA%20\(2002\)%20Vol.3,%20No.1,%2036-45.pdf](http://www.fmipa.its.ac.id/isi%20mipajurnal/jurnal/KAPPA%20(2002)%20Vol.3,%20No.1,%2036-45.pdf)

Model Time Series Box Jenkins Pengguna Transportasi Udara di Bandara Bengkulu

Amaida¹, Sigit Nugroho², dan Jose Rizal²

1. Alumni Jurusan Matematika FMIPA Universitas Bengkulu
2. Staf Pengajar Jurusan Matematika FMIPA Universitas Bengkulu

ABSTRAK

Time series (deret berkala) adalah suatu barisan data yang diamati dan dicatat secara terus menerus. Sebagai contoh pengguna transportasi udara baik yang datang maupun yang berangkat dari dan ke Bandara Bengkulu. Analisis data meliputi plot data *time series*, ACF, PACF, memilih dan menguji model yang sesuai untuk data, serta melakukan peramalan terhadap series untuk beberapa periode kedepan.

Tujuan penelitian ini adalah untuk menerapkan teori analisis deret waktu dalam menentukan model *time series* Box-Jenkins serta melakukan peramalan (*forecasting*) berdasarkan data bulanan selama kurun waktu 5 tahun dari tahun 2002 sampai dengan tahun 2006 dari banyaknya pengguna transportasi udara setiap bulannya di Bandara Bengkulu serta melakukan peramalan (*forecasting*) untuk 3 bulan.

Metode analisis yang digunakan adalah Model *Time Series* Box Jenkins ini ialah melalui tiga tahapan diantaranya identifikasi model, pengestimasi model dan diagnosa model.

Setelah melalui tahap identifikasi model, estimasi model dan diagnosa model maka model yang terbentuk dengan model ARIMA adalah (1,1,0) untuk pengguna transportasi udara yang datang ke Bandara Bengkulu yang dinyatakan dengan persamaan: $Z_t = 1,545 Z_{t-1} - 0,545 Z_{t-2}$, sedangkan model yang terbentuk pada pengguna transportasi udara yang berangkat dari Bandara Bengkulu adalah model ARIMA (1,1,0) yang dinyatakan dengan persamaan: $Z_t = 1,598 Z_{t-1} - 0,598 Z_{t-2}$

Kata kunci: *Time series*, ACF, PACF, ARIMA, Peramalan, Identifikasi Model, Pengestimasi Model dan Diagnosa Model.

PENDAHULUAN

Transportasi udara adalah setiap kegiatan dengan menggunakan pesawat udara untuk menampung penumpang, kargo dan pos untuk satu perjalanan atau lebih dari satu Bandar udara ke Bandar udara lain atau beberapa Bandar udara dalam tempo waktu yang cukup singkat. Kebijakan penambahan penerbangan merupakan salah satu strategi dalam meningkatkan efektifitas penerbangan. Setelah kebijaksanaan ditetapkan, jumlah penumpang yang berangkat dari Bengkulu dan datang ke Bengkulu ternyata mengalami perubahan dibandingkan sebelum penambahan jadwal penerbangan (Badan Pusat Statistik, 2004). Oleh karena tidak dapat dipastikan bagaimana kecenderungan pola musiman, pergerakan naik turun (*fluktuasi*) serta keragaman acak yang terbentuk dari pengguna transportasi udara secara umum, dengan demikian analisis deret waktu dapat digunakan untuk meramalkan jumlah pengguna transportasi udara di Bandara Bengkulu.

Analisis deret waktu digunakan untuk melakukan analisis data dengan mempertimbangkan pengaruh waktu dan data pengamatan yang dikumpulkan secara periodik berdasarkan urutan waktu (jam, hari, minggu, bulan, kuartal, tahun dan sebagainya). Deret waktu (*Time Series*) adalah suatu barisan data yang diamati dan dicatat secara terus menerus. Tujuan analisis deret waktu adalah untuk mendapatkan suatu ukuran yang dapat digunakan untuk membuat keputusan masa kini, untuk prediksi atau peramalan beberapa periode kedepan, dan untuk perencanaan operasional di masa yang akan datang (Clark dan Schkade dalam Ningsih, 2006).

Dalam metode Box Jenkins ini terdapat beberapa model yang akan digunakan diantaranya adalah model *Autoregressive* (AR), model Rata-rata Bergerak (MA), *Autoregressive Moving-Average* (ARMA) serta model ARIMA. Adapun tujuan dari penelitian ini adalah untuk menerapkan teori analisis deret waktu dalam menentukan model *time series* Box-Jenkins serta melakukan peramalan

(forecasting) berdasarkan data bulanan selama kurun waktu 5 tahun dari tahun 2002 sampai dengan tahun 2006 dari banyaknya pengguna transportasi udara setiap bulannya di Bandara B engkulu serta melakukan peramalan (forecasting) untuk 3 bulan mendatang.

MODEL-MODEL TIME SERIES BOX JENKINS

Model Autoregressive (AR)

Model *Autoregressive* yaitu meregresikan atau meramalkan nilai peubah yang tidak diketahui sebagai fungsi nilai-nilai peubah yang telah diketahui.

| Model | Persamaan | ACF | PACF |
|-------|---|---|---|
| AR(1) | $Z_t = \phi Z_{t-1} + a_t$ | $\rho_k = \frac{\gamma_k}{\gamma_0} = \phi^k \quad \text{untuk } k = 1, 2, \dots$ | $\phi_{kk} = \begin{cases} \rho_1 = \phi_1 & k = 1 \\ 0 & k \geq 2 \end{cases}$ |
| AR(2) | $Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + a_t$ | $\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} \quad k \geq 1$ | $\phi_{kk} = \begin{cases} \frac{\phi_1}{1 - \phi_2} & k = 1 \\ \phi_2 & k = 2 \\ 0 & k \geq 3 \end{cases}$ |

Model Moving Average (MA)

Rata-rata bergerak diperoleh melalui pejumlahan dan pencarian nilai rata-rata dari sejumlah periode tertentu, setiap kali menghilangkan nilai yang lama dan menambahkannya dengan nilai yang baru.

| Model | Persamaan | ACF | PACF |
|-------|------------------------------|--|--|
| MA(1) | $Z_t = a_t - \theta a_{t-1}$ | $\rho_k = \frac{\gamma_k}{\gamma_0} = \begin{cases} -\theta & k = 1 \\ 0 & k \geq 2 \end{cases}$ | $\phi_{kk} = \begin{cases} \frac{-\theta(1-\theta^2)}{1-\theta_1^4} & k = 1 \\ \frac{-\theta^2(1-\theta^2)}{(1-\theta^6)} & k = 2 \\ \frac{-\theta_1^3(1-\theta_1^2)}{(1-\theta_1^8)} & k = 3 \end{cases}$ |

| | | | |
|-------|---|--|--|
| MA(2) | $Z_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$ | $\rho_k = \begin{cases} \frac{-\theta_1(1-\theta_2)}{1+\theta_1^2+\theta_2^2} & k=1 \\ \frac{-\theta_2}{1+\theta_1^2+\theta_2^2} & k=2 \\ 0 & k>2 \end{cases}$ | $\rho_k = \begin{cases} \frac{-\theta_1 - \theta_1 \theta_2 + \dots + \theta_2 \theta_1^k}{1+\theta_1^2+\theta_2^2} & k=1, 2, \dots, q \\ 0 & k>q \end{cases}$ |
|-------|---|--|--|

Model ARMA (Autoregressive Moving Average)

| Modell | Persamaan | ACF |
|------------|---|--|
| ARMA (p,q) | $Z_t = \phi_1 Z_{t-1} + \dots + \phi_p Z_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}$ | $\rho_k = \frac{(1-\theta\phi)(\phi-\theta)}{1-2\theta\phi+\theta^2}$ untuk $k \geq 1$ |

Sehingga dari uraian mengenai beberapa model *time series* diatas dapat disimpulkan karakteristik untuk setiap model berdasarkan nilai fungsi autokorelasi (ACF – *Autocorrelation Function*) dan fungsi parsial autokorelasi (PACF – *Partial Autocorrelation Function*), adalah sebagai berikut:

| Proses | ACF | PACF |
|-----------|---|---|
| AR(p) | Menurun secara eksponensial | Cut off pada lag p |
| MA(q) | Cut off pada lag q | Menurun secara eksponensial |
| ARMA(p,q) | Menurun secara eksponensial mulai lag q | Menurun secara eksponensial mulai lag p |

Model ARIMA (Autoregressive Integrated Moving Average)

Model ARIMA merupakan model untuk deret tak stasioner. Suatu deret $\{Z_t\}$ dikatakan mengikuti suatu model ARIMA jika diferensi ke -d, yakni $W_t = \nabla^d Z_t$ merupakan proses ARMA stasioner. Jika W_t adalah ARMA (p,q), maka Z_t dikatakan ARIMA (p,d,q)

Perhatikan proses ARIMA(p,1,q). dengan mendefenisikan $W_t = Z_t + Z_{t-1}$ akan diperoleh:

$$W_t = \phi_1 W_{t-1} + \phi_2 W_{t-2} + \dots + \phi_p W_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_p a_{t-p}$$

atau jika ditulis dalam suku-suku deret pengamatan, akan didapatkan:

$$Z_t - Z_{t-1} = \phi_1 (Z_{t-1} - Z_{t-2}) + \phi_2 (Z_{t-2} - Z_{t-3}) + \dots + \phi_p (Z_{t-p} - Z_{t-p-1}) + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_p a_{t-p}$$

METODE BOX-JENKINS

Identifikasi Model

Langkah ini untuk menentukan apakah data deret waktu yang digunakan bersifat stasioner atau tidak. Jika data deret waktu tersebut tidak stasioner, biasanya dapat digunakan metode perbedaan (*differencing method*). Jika data runtun waktu telah stasioner, maka harus mengidentifikasi bentuk model yang akan digunakan.

Pengestimasi Model

Setelah berhasil menetapkan identifikasi model sementara, selanjutnya parameter-parameter AR dan MA, musiman dan tidak musiman harus ditetapkan dengan cara yang terbaik. Terdapat dua cara yang mendasar untuk mendapatkan parameter-parameter terbaik tersebut yaitu: dengan cara mencoba-coba (*trial and error*) – menguji beberapa nilai yang berbeda dan memilih satu nilai tersebut yang meminimumkan jumlah kuadrat nilai sisa (*sum of squared residuals*) dan perbaikan secara iterative memilih taksiran awal dan kemudian dilakukan penghalusan dengan menggunakan program komputer.

Diagnosa Model

Diagnosa model dapat dilakukan dengan menganalisa plot dari fungsi autokorelasi (*ACF – Autocorrelation Function*) dan fungsi parsial autokorelasi (*PACF – Partial Autocorrelation Function*). Hal ini dikarenakan setiap model mempunyai karakteristik *ACF* dan *PACF* yang berbeda.

Kelayakan suatu model dapat diuji dengan menggunakan uji statistik *Box-Pierce Q*. Uji statistiknya adalah:

$$Q = n \sum_{k=1}^m r_k^2 \quad \text{atau} \quad Q = (N-d) \sum_{k=1}^m r_k^2$$

Statistik *Box-Pierce Q* berdistribusi *Chi-Square* (χ^2) dengan derajat bebas $k-p-q$, dimana k adalah jumlah maksimum lag yang diinginkan, p adalah orde model *autoregressive* sedangkan q adalah orde model *moving average*.

Dengan kriteria pengujian:

Jika $Q \leq \chi_{k-p-q}^2$ atau $p \text{ value} > 0,05$ maka model sesuai (nilai *error* bersifat acak).

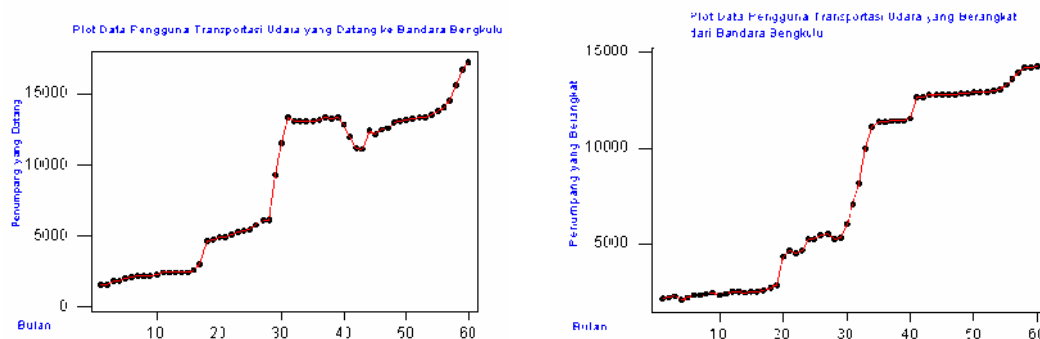
Jika $Q > \chi_{k-p-q}^2$ atau $p \text{ value} < 0,05$ maka model tidak sesuai (nilai *error* tidak bersifat acak).

PENERAPAN TEORI

Untuk menentukan model *time series* yang sesuai dengan data tersebut akan dilakukan langkah-langkah mengidentifikasi model *time series* sebagai berikut:

1. Memplot data terhadap waktu
2. Menguji ada tidaknya trend atau musiman
3. Mengidentifikasi model melalui plot *ACF* dan *PACF* berdasarkan data yang sudah stasioner
4. Menaksir parameter
5. Memprediksi

Identifikasi Model

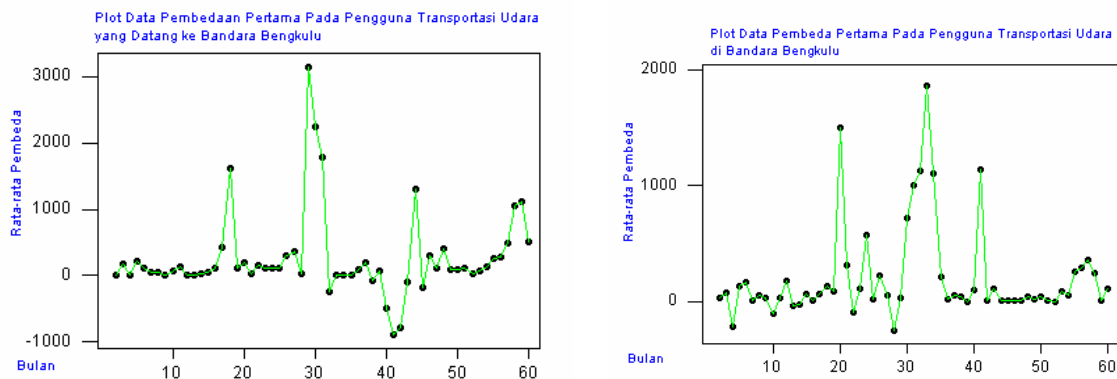


Gambar 1. Plot Data Pengguna Transportasi Udara yang Datang dan berangkat dari dan ke Bandara Bengkulu Mulai Bulan Januari 2002 Sampai Dengan Desember 2006

Plot data di atas baik plot data yang datang maupun yang berangkat dari dan ke Bengkulu menunjukkan adanya kecenderungan pergerakan naik turun (*fluktuasi*) dari tahun ke tahun. Namun bukan hanya terjadi peningkatan secara umum tapi terlihat pula bahwa variasi kenaikannya pun beragam. Pada tahun 2002 terlihat kenaikan yang terjadi tidak begitu tinggi, keadaan ini disebabkan oleh masih mahalnnya harga tiket dari dan ke Bengkulu, kuantitas jadwal penerbangan yang masih sedikit baik yang berangkat maupun yang datang dari dan ke Bengkulu, serta pesawat terbang yang beroperasi pada tahun tersebut hanya Merpati Air Lines. Akan tetapi pada pertengahan tahun 2003 sampai dengan tahun 2006 jumlah pengguna transportasi di Bandara Bengkulu mengalami kenaikan yang cukup tinggi ini terlihat dari variasi yang begitu besar, kenaikan jumlah pengguna transportasi Udara di Bandara Bengkulu ini didukung dengan harga tiket sudah relatif murah, banyaknya jadwal penerbangan yang berangkat maupun yang datang dari dan ke Bengkulu dalam seminggu, serta pesawat terbang yang beroperasi sudah cukup banyak diantaranya: Sriwijaya Air, Batavia Air, Adam Air, Lion Air dan Riau Air Lines.

Pembedaan (*Differensial*) Pertama

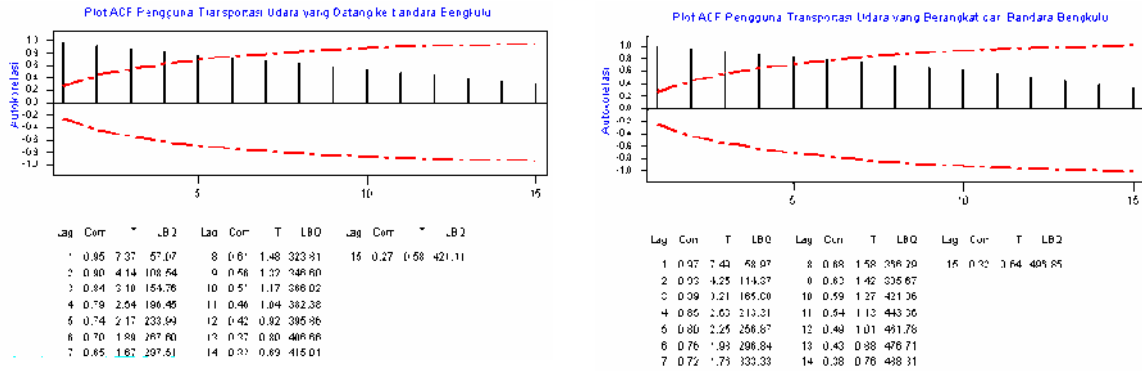
Pembedaan (*Differensial*) pertama ini dilakukan jika plot data yang diperoleh tidak stasioner artinya varian serta rata-ratanya tidak berada dalam keadaan konstan. Seperti yang sudah terlihat plot data pengguna transportasi udara di Bandara Bengkulu baik yang datang maupun yang berangkat diatas bahwa plot datanya tidak stasioner, untuk itu untuk menstasionerkan data tersebut dilakukan pembedaan pertama pada data aslinya.



Gambar 2. Plot Data Pembedaan Pertama Pada Pengguna Transportasi Udara yang Datang dan Berangkat dari dan ke Bandara Bengkulu.

Autokorelasi (*Autocorrelation*)

Setelah dilakukan pembedaan pertama, terlihat bahwa plot grafik yang terbentuk tampak plot data diatas menunjukkan pola yang stasioner. Berikut ini adalah plot ACF dari data yang sudah diturunkan satu lag.



Gambar 3. Plot ACF Data Pengguna Transportasi Udara yang Datang dan Berangkat dari dan ke Bandara Bengkulu

Plot autokorelasi diatas baik plot autokorelasi yang datang maupun yang berangkat menunjukkan 15 lag. Ini diperoleh dari lag sebanyak $n/4$ untuk pengamatan (n) ≤ 240 (Iriawan, N dan Astuti, P. S., 2006). Dalam kasus ini, jumlah pengamatan pengguna transportasi udara di Bandara Bengkulu adalah 60 sehingga $60/4 = 15$. Oleh karena itu, MINITAB menggambarkan 15 lag. Dalam kasus ini, selang kepercayaan yang digunakan adalah $\alpha = 5\%$. Dengan menggunakan tabel distribusi normal, nilai z yang hampir mendekati $p=5\%$ adalah 1.645. Oleh karena itu, nilai z yang dijadikan sebagai patokan adalah statistik $z_{0,05}=1,645$. Statistik T ini digunakan untuk uji hipotesis yang mendeteksi korelasi pada lag ke z_t dengan z_{t-k} (k adalah lag). Dalam hal ini, garis merah adalah selang kepercayaan yang merupakan garis batas signifikan autokorelasi. Kemudian output menunjukkan bahwa nilai-nilai autokorelasi yang turun secara eksponensial pada nilai autokorelasi positif. Pola seperti ini dapat dijadikan petunjuk awal mendeteksi model *time series*.

Prosedur pengujian hipotesis secara statistik terdiri dari beberapa langkah. Langkah-langkah tersebut adalah:

1. Merumuskan Hipotesis

$H_0 : \rho_k = 0$ menjelaskan bahwa antara z_t dengan z_{t-k} tidak ada korelasi

$H_1 : \rho_k \neq 0$ menjelaskan bahwa antara z_t dengan z_{t-k} ada korelasi

2. Menentukan Taraf Signifikan

Taraf signifikan yang digunakan yaitu 5%.

3. Menentukan Uji Statistik

Uji statistik adalah suatu nilai yang diperoleh dari sampel dan digunakan untuk memutuskan atau menolak hipotesa. Statistik uji yang digunakan adalah statistik uji T.

4. Mengambil Keputusan atau Kesimpulan

Jika statistik $T > z_{0,05}$, maka tolak hipotesis awal (H_0)

Jika statistik $T < z_{0,05}$, maka terima hipotesis awal (H_0)

Output pada plot autokorelasi pengguna transportasi udara yang berangkat dari Bandara Bengkulu menunjukkan nilai statistik T setelah lag k yang melebihi nilai statistik $z_{0,05}$ yaitu pada :

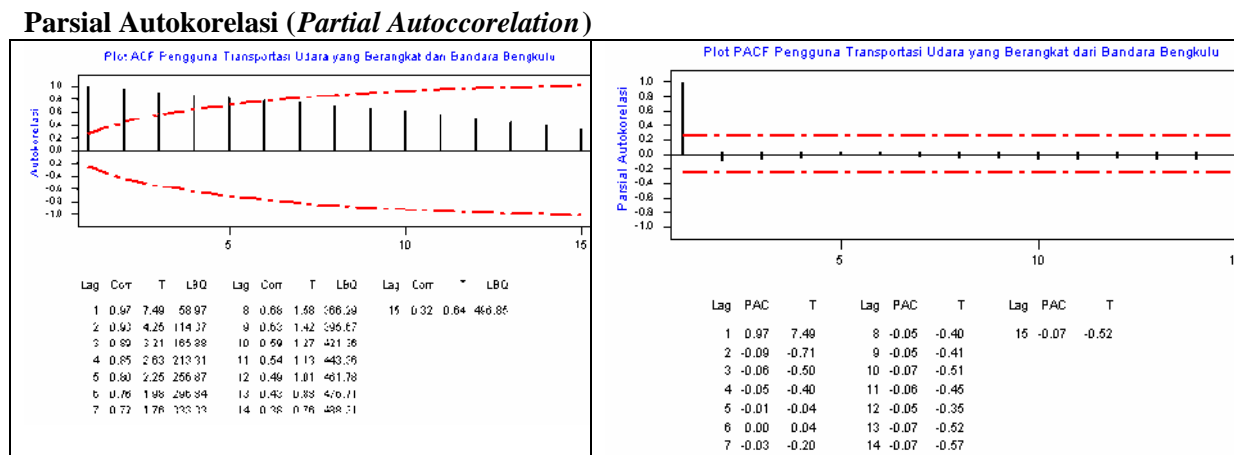
1. Lag 1 bernilai 7,49
2. Lag 2 bernilai 4,25
3. Lag 3 bernilai 3,21
4. Lag 4 bernilai 2,63
5. Lag 5 bernilai 2,25
6. Lag 6 bernilai 1,98

Sedangkan output pada plot autokorelasi pengguna transportasi udara yang datang ke Bandara Bengkulu menunjukkan nilai statistik T setelah lag k yang melebihi nilai statistik $z_{0,05}$ yaitu pada :

1. Lag 1 bernilai 7,37

2. Lag 2 bernilai 4,14
3. Lag 3 bernilai 3,10
4. Lag 4 bernilai 2,54
5. Lag 5 bernilai 2,17
6. Lag 6 bernilai 1,89

Dari gambar plot ACF data pengguna transportasi udara baik yang datang maupun berangkat dari dan ke Bandara Bengkulu terlihat bahwa pada lag 1, lag 2, lag 3, lag 4, lag 5 dan lag 6 menunjukkan lokasi statistik T jatuh didaerah penolak an. Sehingga pada lag 1, lag 2, lag 3, lag 4, lag 5 dan lag 6 atau antara z_t dengan z_{t-k} secara statistik ada korelasi ($\rho_k \neq 0$) cukup berarti. Selain di lag 1, lag 2, lag 3, lag 4, lag 5 dan lag 6, korelasinya secara statistik tidak berarti ($\rho_k = 0$).



Gambar 4. Plot PACF Data Pengguna Transportasi Udara yang Berangkat dari Bandara Bengkulu

PACF pada gambar 2. menunjukkan bahwa setelah lag 1, PACF mengalami *cut off*. Disamping itu, pada lag 1, PACF tampak keluar dari selang kepercayaan. Berdasarkan pada karakteristik yang dimiliki oleh masing-masing model di mana ACF turun secara eksponensial dan PACF turun setelah lag 1, ini berarti ada indikasi modelnya adalah model *autoregressive*. Berdasarkan hasil, dapat diperkirakan bahwa model *time series*-nya adalah model *autoregressive*. Karena PACF turun pada lag 1, maka perkiraan model *autoregressive* berorde 1 atau AR(1).

Pengstimasian Model

Dengan menggunakan program MINITAB, dibawah ini merupakan taksiran parameter dari model AR(1).

ARIMA model for Datang

| Estimates at each iteration | | | |
|-----------------------------|----------|------------|---------|
| Iteration | SSE | Parameters | |
| 0 | 22230104 | 0,100 | 338,559 |
| 1 | 20117764 | 0,539 | 399,328 |
| 2 | 19090689 | 0,543 | 459,817 |
| 3 | 18982855 | 0,544 | 462,604 |
| 4 | 18982493 | 0,545 | 495,172 |

ARIMA model for Berangkat

| Estimates at each iteration | | | |
|-----------------------------|---------|------------|---------|
| Iteration | SSE | Parameters | |
| 0 | 8890082 | 0.100 | 186.161 |
| 1 | 7959838 | 0.250 | 154.799 |
| 2 | 7466617 | 0.400 | 123.274 |
| 3 | 7380513 | 0.489 | 104.093 |
| 4 | 7380208 | 0.494 | 102.469 |

Model Time Series Box-Jenkins Pengguna Transportasi Udara ...

| | | | | | | | |
|---|----------|---------|---------|---|---------|---------|---------|
| 5 | 18982492 | 0,545 | 496,061 | 5 | 7380207 | 0.494 | 102.345 |
| Relative change in each estimate less than 0,0010 | | | | Relative change in each estimate less than 0.0010 | | | |
| Final Estimates of Parameters | | | | Final Estimates of Parameters | | | |
| Type | Coef | SE Coef | T | Type | Coef | SE Coef | T |
| P | | | | P | | | |
| AR 1 | 0,545 | 0,1173 | 4,96 | AR 1 | 0.4943 | 0.1152 | 4.29 |
| 0,000 | | | | 0.000 | | | |
| Constant | 141,06 | 75,11 | 1,88 | Constant | 102.35 | 46.83 | 2.19 |
| 0,065 | | | | 0.033 | | | |
| Differencing: 1 regular difference | | | | Differencing: 1 regular difference | | | |
| Number of observations: Original series 60, after differencing 59 | | | | Number of observations: Original series 60, after differencing 59 | | | |
| Residuals: SS = 18970910 (backforecasts excluded) | | | | Residuals: SS = 7374973 (backforecasts excluded) | | | |
| MS = 345146 DF = 57 | | | | MS = 129385 DF = 57 | | | |

Gambar 4. Output Data Pengguna Transportasi Udara yang Datang dan Berangkat dari dan ke Bandara Bengkulu

Tahap awal menginterpretasikan output *time series* adalah melihat signifikansi parameter model yang telah dimodelkan. Seperti yang telah dijelaskan diatas, bahwa model *time series* yang digunakan adalah model AR(1) yang melibatkan konstanta didalamnya. Jadi ada 2 parameter dalam model. Secara umum, signifikansi konstanta tidak perlu diuji sehingga disini yang diuji hanya parameter *autoregressive* (ϕ).

Prosedur pengujian hipotesis secara statistik terdiri dari beberapa langkah. Langkah-langkah tersebut adalah:

1. Merumuskan Hipotesis

$H_0 : \phi = 0$ menjelaskan bahwa parameter AR tidak cukup signifikan dalam model

$H_1 : \phi \neq 0$ menjelaskan bahwa parameter AR cukup signifikan dalam model

2. Menentukan Taraf Signifikan

Taraf signifikan yang digunakan yaitu 5%.

3. Menentukan Uji Statistik

Uji statistik adalah suatu nilai yang diperoleh dari sampel dan digunakan untuk memutuskan atau menolak hipotesa. Statistik uji yang digunakan adalah statistik uji T.

4. Mengambil Keputusan atau Kesimpulan

Jika statistik $T > z_{0,05}$, maka tolak hipotesis awal (H_0)

Jika statistik $T < z_{0,05}$, maka terima hipotesis awal (H_0)

Berdasarkan tabel taksiran parameter, hasil pengolahan data yang ditunjukkan statistik T untuk parameter AR(1) atau ϕ adalah 3,92 (Pengguna transportasi udara yang Datang) sedangkan ϕ adalah 5,68 (Pengguna Transportasi yang Berangkat). Pada $\alpha = 5\%$, statistik z adalah 1.645. Bila statistik T dibandingkan dengan nilai z , maka nilai statistik T lebih besar. Berdasarkan informasi, maka dapat disimpulkan bahwa dalam model AR(1), parameter ϕ bagi pengguna transportasi udara yang datang maupun yang berangkat dari dan ke Bandara Bengkulu cukup signifikan. Jadi, persamaan modelnya adalah:

| Model untuk Pengguna transportasi udara yang datang ke Bandara Bengkulu | Model untuk Pengguna transportasi udara yang berangkat dari Bandara Bengkulu |
|---|---|
| $Z_t - Z_{t-1} = \phi(Z_{t-1} - Z_{t-2})$ $Z_t - Z_{t-1} = 0,4669(Z_{t-1} - Z_{t-2})$ $Z_t = Z_{t-1} + 0,4669Z_{t-1} - 0,4669Z_{t-2}$ $Z_t = Z_{t-1}(1 + 0,4669) - 0,4669Z_{t-2}$ $Z_t = 1,4669Z_{t-1} - 0,4669Z_{t-2}$ | $Z_t - Z_{t-1} = \phi(Z_{t-1} - Z_{t-2})$ $Z_t - Z_{t-1} = 0,4943(Z_{t-1} - Z_{t-2})$ $Z_t = Z_{t-1} + 0,4943Z_{t-1} - 0,4943Z_{t-2}$ $Z_t = Z_{t-1}(1 + 0,4943) - 0,4943Z_{t-2}$ $Z_t = 1,4943Z_{t-1} - 0,4943Z_{t-2}$ |

Diagnosa Model

Untuk mendeteksi adanya proses *white noise*, akan dilakukan beberapa uji. Uji pertama adalah uji korelasi yang berguna untuk mendeteksi independensi residual dan uji kedua adalah kemormalan residual model.

| Statistik Ljung Box Pierce untuk Pengguna transportasi udara yang datang ke Bandara Bengkulu | Statistik Ljung Box Pierce untuk Pengguna transportasi udara yang berangkat dari Bandara Bengkulu | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|--|---|-------|-------|-------|--|-----|----|----|----|----|------------|-----|------|------|------|----|----|----|----|----|---------|-------|-------|-------|-------|--|--|--|--|--|--|-----|----|----|----|----|------------|-----|------|------|------|----|----|----|----|----|---------|-------|-------|-------|-------|
| <table style="width: 100%; border-collapse: collapse;"> <tr> <td style="text-align: left;">Modified Box-Pierce (Ljung-Box) Chi-Square statistic</td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Lag</td> <td style="text-align: center;">12</td> <td style="text-align: center;">24</td> <td style="text-align: center;">36</td> <td style="text-align: center;">48</td> </tr> <tr> <td>Chi-Square</td> <td style="text-align: center;">4.9</td> <td style="text-align: center;">19.2</td> <td style="text-align: center;">21.3</td> <td style="text-align: center;">22.4</td> </tr> <tr> <td>DF</td> <td style="text-align: center;">10</td> <td style="text-align: center;">22</td> <td style="text-align: center;">34</td> <td style="text-align: center;">46</td> </tr> <tr> <td>P-Value</td> <td style="text-align: center;">0.897</td> <td style="text-align: center;">0.630</td> <td style="text-align: center;">0.956</td> <td style="text-align: center;">0.999</td> </tr> </table> | Modified Box-Pierce (Ljung-Box) Chi-Square statistic | | | | | Lag | 12 | 24 | 36 | 48 | Chi-Square | 4.9 | 19.2 | 21.3 | 22.4 | DF | 10 | 22 | 34 | 46 | P-Value | 0.897 | 0.630 | 0.956 | 0.999 | <table style="width: 100%; border-collapse: collapse;"> <tr> <td style="text-align: left;">Modified Box-Pierce (Ljung-Box) Chi-Square statistic</td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Lag</td> <td style="text-align: center;">12</td> <td style="text-align: center;">24</td> <td style="text-align: center;">36</td> <td style="text-align: center;">48</td> </tr> <tr> <td>Chi-Square</td> <td style="text-align: center;">4.9</td> <td style="text-align: center;">19.2</td> <td style="text-align: center;">21.3</td> <td style="text-align: center;">22.4</td> </tr> <tr> <td>DF</td> <td style="text-align: center;">10</td> <td style="text-align: center;">22</td> <td style="text-align: center;">34</td> <td style="text-align: center;">46</td> </tr> <tr> <td>P-Value</td> <td style="text-align: center;">0.897</td> <td style="text-align: center;">0.630</td> <td style="text-align: center;">0.956</td> <td style="text-align: center;">0.999</td> </tr> </table> | Modified Box-Pierce (Ljung-Box) Chi-Square statistic | | | | | Lag | 12 | 24 | 36 | 48 | Chi-Square | 4.9 | 19.2 | 21.3 | 22.4 | DF | 10 | 22 | 34 | 46 | P-Value | 0.897 | 0.630 | 0.956 | 0.999 |
| Modified Box-Pierce (Ljung-Box) Chi-Square statistic | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Lag | 12 | 24 | 36 | 48 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Chi-Square | 4.9 | 19.2 | 21.3 | 22.4 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| DF | 10 | 22 | 34 | 46 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| P-Value | 0.897 | 0.630 | 0.956 | 0.999 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Modified Box-Pierce (Ljung-Box) Chi-Square statistic | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Lag | 12 | 24 | 36 | 48 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Chi-Square | 4.9 | 19.2 | 21.3 | 22.4 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| DF | 10 | 22 | 34 | 46 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| P-Value | 0.897 | 0.630 | 0.956 | 0.999 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Uji Independensi Residual

Uji ini dilakukan untuk mendeteksi independensi residual antar-lag. Dua lag dikatakan tidak berkorelasi apabila antar-lag tidak ada korelasi cukup berarti. Dalam *time series*, uji dilakukan dengan menggunakan statistik Ljung-Box-Pierce.

1. Pengujian Hipotesis

Untuk mendeteksi adanya proses idependensi antar-residual, hipotesisnya adalah:

$$H_0 : \rho_{a_t, a_{t+k}} = 0 \text{ (ada korelasi antar-lag)}$$

$$H_1 : \rho_{a_t, a_{t+k}} \neq 0 \text{ (tidak ada korelasi antar-lag)}$$

2. Menentukan Taraf Signifikan

Taraf signifikan yang digunakan adalah $\alpha = 5\%$

3. Uji Statistik

$$\text{Uji Statistik yang digunakan Ljung Pierce } Q = n \sum_{k=1}^m r_k^2 \text{ atau } Q = (N - d) \sum_{k=1}^m r_k^2$$

4. Pengambilan keputusan

$$\text{Jika } Q \leq \chi_{k-p-q}^2 = p \text{ value} > 0,05 \text{ maka terima } H_0.$$

$$\text{Jika } Q > \chi_{k-p-q}^2 = p \text{ value} < 0,05 \text{ maka tolak } H_0.$$

Pada tahap ini akan digunakan level teloransi (α) sebesar 5% untuk analisis. Deteksi idependensi antar-lag dilakukan pada tiap lag. Tabel 1.a menampilkan nilai statistik Ljung-Box-Pierce pada lag 12, 24, 36, dan 48. Nilai statistik Ljung-Box-Pierce pada lag 12 berarti menunjukkan nilai statistik Ljung-Box-Pierce antara lag t dengan lag 12. untuk mengevaluasi output, lihat tabel 1.a dibawah ini:

Tabel 1.a.

Pengguna Transportasi Udara yang Datang ke Bandara Bengkulu

| Lag (K) | df ($K-k$) | Statistik Ljung-Box-Pierce. | χ^2 tabel | p-value |
|-------------|--------------|-----------------------------|----------------|---------|
| 12 | 10 (12-2) | 8,1 | 18,31 | 0,620 |
| 24 | 22 (24-2) | 20,7 | 33,92 | 0,540 |
| 36 | 34 (36-2) | 28,9 | 48,602 | 0,716 |
| 48 | 46 (48-2) | 30,0 | | 0,968 |

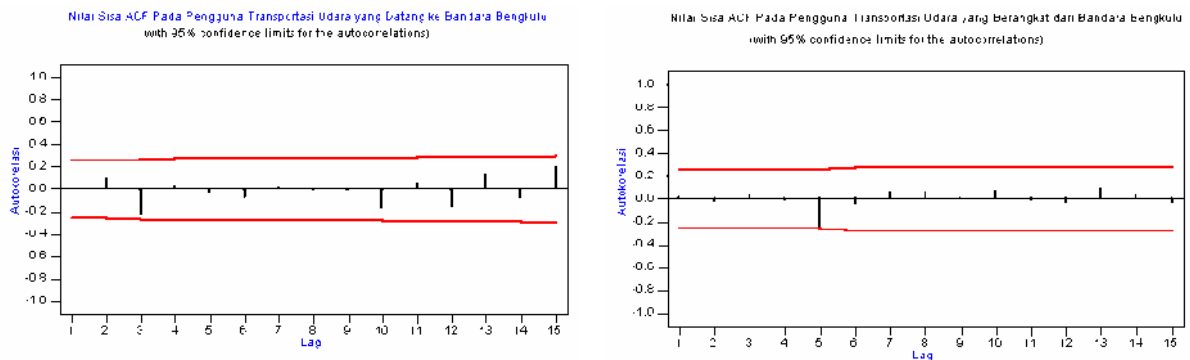
Tabel 1.b.

Pengguna Transportasi Udara yang Berangkat dari Bandara Bengkulu

| Lag (K) | df ($K-k$) | Statistik Ljung-Box-Pierce. | χ^2 tabel | p-value |
|-------------|--------------|-----------------------------|----------------|---------|
| 12 | 10 (12-2) | 4,9 | 18,31 | 0,897 |
| 24 | 22 (24-2) | 19,2 | 33,92 | 0,630 |
| 36 | 34 (36-2) | 21,3 | 48,602 | 0,956 |
| 48 | 46 (48-2) | 22,4 | | 0,999 |

Sehingga dari data Pengguna transportasi udara yang datang ke Bandara Bengkulu dan Pengguna transportasi udara yang berangkat dari Bandara Bengkulu dapat disimpulkan bahwa tidak ada korelasi antara residual pada lag t dengan residual pada lag 12 karena statistik Ljung -Box-Pierce tidak lebih dari $\chi^2_{(5\%,10)}$. Begitu pula untuk lag 24, 36 dan 48, statistik Ljung -Box-Pierce tidak melebihi $\chi^2_{(5\%,22)}$, $\chi^2_{(5\%,36)}$ dan $\chi^2_{(5\%,48)}$. Artinya, antara residual pada lag t dengan residual pada (sampai) lag 48, tidak ada yang saling berkorelasi. Sehingga dapat disimpulkan bahwa residual telah memenuhi asumsi independensi.

Uji Kenormalan Residual



Gambar 5. Plot ACF Residual Pengguna Transportasi Udara yang Datang maupun yang Berangkat dari dan ke Bandara Bengkulu

Uji lain yang dapat digunakan untuk melihat independensi antar -lag yaitu dengan membuat plot ACF residual seperti pada gambar 2 diatas. Dengan memperhatikan gambar diatas terlihat bahwa tidak ada 1 lag pun yang keluar batas. Berdasarkan grafik ACF, dapat disimpulkan bahwa residual model telah independen. Hasil uji Ljung Box dan Plot ACF residual konsisten.

PENUTUP

Kesimpulan

Dari hasil penelitian yang diperoleh, dapat ditarik beberapa kesimpulan sebagai berikut:

1. Berdasarkan pada grafik data dari pengguna transportasi udara yang datang maupun yang berangkat dari dan ke Bandara Bengkulu selama periode tahun 2002 sampai dengan tahun 2006 belum stasioner pada level awal, sehingga perlu dilakukan proses pembedaan (*differencing*) hingga data stasioner, pada pembedaan pertama ternyata kedua data sudah stasioner sehingga tidak perlu dilakukan pembedaan (*differencing*) lagi.
2. Model terbaik yang terbentuk dari pengguna transportasi udara yang datang ke Bandara Bengkulu adalah ARIMA (1,1,0) yang dinyatakan dengan persamaan:

$$Z_t = 1,545 Z_{t-1} - 0,545 Z_{t-2}$$

3. Model terbaik yang terbentuk dari pengguna transportasi udara yang datang ke Bandara Bengkulu adalah ARIMA (1,1,0) yang dinyatakan dengan persamaan:

$$Z_t = 1,598 Z_{t-1} - 0,598 Z_{t-2}$$

4. Hasil peramalan (*forecasting*) bagi pengguna transportasi udara yang datang ke Bandara Bengkulu selama tiga bulan kedepan, menunjukkan peningkatan untuk setiap bulannya yaitu pada bulan Januari sebesar 14373 orang, bulan Februari sebesar 14411 orang dan bulan Maret sebesar 14434 orang. Sedangkan untuk hasil peramalan peramalan (*forecasting*) bagi pengguna transportasi udara yang berangkat dari Bandara Bengkulu selama tiga bulan kedepan juga menunjukkan peningkatan untuk setiap bulannya yaitu pada bulan Januari sebesar 17496 orang, bulan Februari sebesar 17646 orang dan bulan Maret sebesar 17727 orang.

Saran

Beberapa saran yang dapat diberikan oleh penulis sebagai bahan penelitian lanjutan adalah sebagai berikut: dalam penentuan nilai p , d , dan q pada model ARIMA diperlukan kecermatan dan ketelitian, sehingga meskipun telah dikembangkan beberapa pedoman yang bermanfaat dalam membantu mempermudah penentuan model tersebut, masih tetap diperlukan pengalaman dan usaha mencoba-coba (*trial and error*) agar dapat menggunakan metode tersebut dengan lebih tepat dan peramalan dengan menggunakan metode ARIMA merupakan alternatif dalam peramalan jumlah pengguna transportasi udara di Bandara Bengkulu dengan menggunakan data historis, sehingga tidak tertutup kemungkinan untuk dilakukan peramalan dengan metode lainnya. Untuk penelitian lebih lanjut diharapkan analisis peramalan tidak hanya pada masalah *time series* saja, tetapi juga peramalan kausal (sebab akibat), sehingga bisa dicari sebab-sebab terjadinya perubahan jumlah pengguna transportasi udara atau variabel-variabel apa yang mempengaruhi jumlah pengguna transportasi udara tersebut, sehingga analisis ini menjadi lebih lengkap.

LAMPIRAN

Lampiran 1. Data Jumlah Pengguna Transportasi Udara yang Datang ke Bandara Bengkulu dari bulan Januari 2002 Sampai Dengan Desember 2006

| Tahun | Bulan | | | | | | | | | | | |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| | Jan | Feb | Mar | Apr | Mei | Juni | Juli | Agst | Sept | Okt | Nov | Des |
| 2002 | 1590 | 1594 | 1771 | 1780 | 1990 | 2102 | 2147 | 2190 | 2193 | 2265 | 2386 | 2387 |
| 2003 | 2389 | 2406 | 2457 | 2575 | 3001 | 4621 | 4723 | 4924 | 4949 | 5105 | 5223 | 5341 |
| 2004 | 5460 | 5768 | 6126 | 6145 | 9302 | 11550 | 13344 | 13098 | 13102 | 13107 | 13109 | 13205 |
| 2005 | 13401 | 13332 | 13407 | 12900 | 11998 | 11200 | 11099 | 12403 | 12209 | 12507 | 13613 | 13020 |
| 2006 | 13119 | 13201 | 13320 | 13336 | 13405 | 13534 | 13790 | 14064 | 14559 | 15607 | 16721 | 17223 |

Sumber : Dinas Perhubungan Propinsi Bengkulu

Lampiran 2. Data Jumlah Pengguna Transportasi Udara yang Berangkat dari Bandara Bengkulu dari bulan Januari 2002 Sampai Dengan Desember 2006

| Tahun | Bulan | | | | | | | | | | | |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| | Jan | Feb | Mar | Apr | Mei | Juni | Juli | Agst | Sept | Okt | Nov | Des |
| 2002 | 2112 | 2146 | 2221 | 2001 | 2136 | 2301 | 2311 | 2368 | 2399 | 2296 | 2324 | 2499 |
| 2003 | 2456 | 2434 | 2501 | 2503 | 2564 | 2698 | 2788 | 4299 | 4599 | 4507 | 4621 | 5199 |
| 2004 | 5223 | 5445 | 5502 | 5245 | 5278 | 6004 | 7012 | 8141 | 10010 | 11115 | 11322 | 11342 |
| 2005 | 11400 | 11438 | 11439 | 11536 | 12678 | 12680 | 12789 | 12796 | 12801 | 12809 | 12811 | 12852 |
| 2006 | 12867 | 12903 | 12938 | 12909 | 13001 | 13056 | 13307 | 13602 | 13956 | 14200 | 14204 | 14301 |

Sumber : Dinas Perhubungan Propinsi Bengkulu

DAFTAR PUSTAKA

- [1]. Anonim, 2004. Statistik Perhubungan Propinsi Bengkulu . Bengkulu: BPS Bengkulu.
- [2]. Anonim, 2000. *Time Series Analysis*. http://www.xycoon.com/arma_II_process.htm. 17 Februari 2007. Pukul 19.48 WIB.
- [3]. Anonim, 2005. Metodologi Penelitian Analisis Time Series . Bandung: Lembaga Penelitian dan Pengabdian Kepada Masyarakat .
- [4]. Arsyad, L. 1995. *Peramalan Bisnis Edisi Pertama*. Yogyakarta: Fakultas Ekonomi, Universitas Gadjah Mada.
- [5]. Assauri, S. 1984. *Teknik dan Metode Peramalan Edisi Pertama Dalam Ekonomi dan Dunia Usaha*. Jakarta: Lembaga Penerbit Fakultas Ekonomi, Universitas Indonesia.
- [6]. Box, G.E.P. and G.M. Jenkins. 1997. *Time Series Analysis : Forecasting and Control*, San Francisco: Holden-Day.
- [7]. Cryer, J. D. 1986. *Time Series Analysis*. Boston : PWS-KENT Publishing Company.
- [8]. Lerbin, R.A. 2002. *Peramalan Bisnis*. Jakarta: Ghalia Indonesia, Jakarta.
- [9]. Manurung, A. 1989. *Ramalan Produksi Kelapa Sawit dengan Model ARIMA* Buletin Perkebunan Jember. Bull 20 (4) : 197 - 217.
- [10]. Nengsih, Y.A. 2006. *Dekomposisi Komponen-Komponen Deret Waktu untuk Peramalan Jumlah Kedatangan Tamu Asing di Indonesia* . Tugas Akhir Sarjana Matematika . Bengkulu: Universitas Bengkulu: (Tidak Dipublikasikan)
- [11]. Situngkir, H. dan Y. Sxxx. 2003. *Peramalan Jangka Pendek Deret Waktu Keuangan di Indonesia*. <http://www.bandungfe.scripsterz.org>. 6 Februari 2007. Pukul 09.32 WIB.
- [12]. Sugiarto dan Harijono. 2000. *Peramalan Bisnis*. Jakarta: Gramedia Pustaka Utama, Jakarta.
- [13]. Ulfa, A. dan A. Yulian. 2003. *Model Alternatif Forecasting Deviden BUMN*. http://www.time-series.com/kajian_Alusin-2.htm. 13 Januari 2007. Pukul 08.58 WIB.

- [14]. Wei, S.W. 1990. *Time Series Analysis*. Departement of Statistics Temple University: Addison - Wesley Publishing Company, Inc.

Analisis Jalur Faktor – Faktor Yang Mempengaruhi Produksi Kacang Tanah Pada Sistem Olah Tanah dan Tanpa Olah Tanah

Yulianti¹⁾, Sigit Nugroho²⁾, Baki Swita²⁾

¹⁾Alumni Jurusan Matematika FMIPA Universitas Bengkulu

²⁾Staf Pengajar Jurusan Matematika FMIPA Universitas Bengkulu

Abstrak

Kacang tanah (*Arachis hypogaea L.*) merupakan salah satu tanaman palawija yang dibudidayakan oleh petani sebagai tanaman produksi. Usahatani kacang tanah di Desa Retak Mudik dilakukan dengan dua cara yaitu dengan mengolah tanah dan tanpa mengolah tanah terlebih dahulu sebelum dilakukan penanaman. Penelitian ini bertujuan untuk mengetahui koefisien jalur faktor-faktor produksi yang dicurahkan petani terhadap hasil produksi tanaman kacang tanah pada Sistem OT dan TOT berdasarkan hasil dari analisis jalur. Serta untuk mengetahui sistem tanam yang terbaik untuk meningkatkan produksi usahatani kacang tanah. Hasil analisis Jalur menunjukkan bahwa pada Sistem OT, variabel yang berpengaruh secara signifikan terhadap jumlah produksi adalah Jumlah Tenaga Kerja, Jumlah Pupuk TSP, dan Jumlah Benih. Pada Sistem TOT, variabel yang berpengaruh secara signifikan terhadap Jumlah Produksi adalah Jumlah Tenaga Kerja, Jumlah Pupuk TSP, Jumlah Pupuk KCl, dan Jumlah Benih. Berdasarkan koefisien jalur yang diperoleh, dapat dikatakan bahwa usahatani kacang tanah pada Sistem OT lebih baik dibandingkan pada Sistem TOT di Desa Retak Mudik.

Kata kunci : Faktor Produksi, Analisis Jalur

PENDAHULUAN

Untuk meningkatkan produksi usahatani kacang tanah, diperlukan faktor-faktor produksi (input) seperti luas lahan, modal (benih, pupuk, pestisida), dan tenaga kerja yang diikuti dengan cara pengolahan tanah dan bercocok tanam yang baik. Oleh karena itu harus diketahui pengaruh setiap faktor produksi terhadap hasil produksi kacang tanah baik secara langsung maupun secara tidak langsung untuk kedua sistem pengolahan tanah. Metode statistika yang dapat digunakan untuk mengetahui hal tersebut adalah analisis jalur (*path analysis*). Dari hasil analisis ini juga dapat dilihat sistem pengolahan tanah yang lebih meningkatkan produksi kacang tanah.

Analisis jalur (*path analysis*) dikembangkan sebagai metode untuk mempelajari pengaruh (efek) secara langsung maupun secara tidak langsung dari variabel bebas terhadap variabel tak bebas. Di dalam melakukan analisis jalur, usaha untuk membangun diagram jalur (*path diagram*) tidak dapat dilepaskan karena dapat memperjelas uraian-uraian yang dikemukakan.

Sehingga dalam kaitannya dengan fakta di atas maka penelitian ini bertujuan untuk mengetahui koefisien jalur faktor-faktor produksi yang dicurahkan petani terhadap hasil produksi tanaman kacang tanah pada sistem OT dan TOT berdasarkan hasil dari analisis jalur serta untuk mengetahui sistem tanam yang terbaik untuk meningkatkan produksi usahatani kacang tanah.

TINJAUAN PUSTAKA

Analisis Jalur

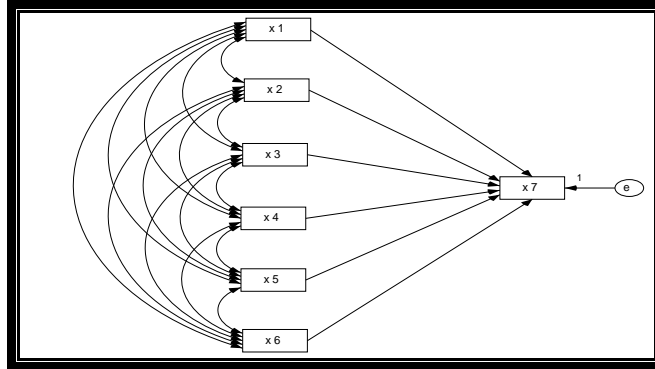
Analisis jalur dan regresi berganda saling berhubungan [6]. Analisis jalur merupakan perluasan dari model regresi [1]. Analisis jalur tidak hanya mengetahui berapa besarnya pengaruh namun juga variabel mana yang merupakan pengaruh langsung atau tidak langsung [3].

Persyaratan atau asumsi yang harus dipenuhi dalam analisis jalur adalah

1. Hubungan antara variabel bebas dan variabel tak bebas di dalam model adalah linier.
2. Variabel yang diamati mempunyai sifat *aditif*.
3. Variabel sisa tidak berkorelasi dengan variabel yang sesudahnya.
4. Variabel yang diukur berskala interval atau rasio.

Model jalur (*Path Model*) adalah suatu diagram hubungan variabel bebas, perantara, dan variabel tak bebas. Diagram jalur, secara grafis sangat membantu untuk melukiskan pola hubungan kausal antara sejumlah variabel [10].

Dalam penelitian ini, diagram jalurnya adalah :



Gambar 1 Diagram Jalur dengan 6 Variabel Bebas

Koefisien jalur adalah koefisien regresi baku yang menunjukkan pengaruh langsung variabel bebas terhadap variabel tak bebas dalam *path model* [2]. Notasi yang digunakan untuk koefisien jalur adalah P_{ij} yang berarti pengaruh variabel j terhadap variabel i .

Konsep koefisien jalur dapat digambarkan sebagai berikut [11]. Misalkan ada m variabel bebas, yakni X_1, X_2, \dots, X_m dan X_0 sebagai variabel tak bebas. X_u merupakan notasi untuk variabel sisa (*residual*) dan semua variabel saling berkorelasi kecuali variabel sisa. Koefisien C_{0i} ($i = 1, 2, \dots, m$) menunjukkan sumbangan nyata X_i secara langsung terhadap X_0 serta semua hubungan adalah linier, maka persamaannya adalah

$$X_0 = C_{01}X_1 + C_{02}X_2 + \dots + C_{0m}X_m + C_{0u}X_u \quad (1)$$

$$\text{Jika } X_i = \frac{\sigma_i}{\sigma_0} V_i \quad \forall i, i = 0, 1, 2, \dots, m$$

keterangan :

V_i = variabel ke- i ($i = 0, 1, 2, \dots, m$)

σ_i = simpangan baku populasi dari variabel bebas ke- i

σ_0 = simpangan baku populasi dari variabel tak bebas

maka persamaan (1) menjadi

$$V_0 = C_{01} \left(\frac{\sigma_1}{\sigma_0} \right) V_1 + C_{02} \left(\frac{\sigma_2}{\sigma_0} \right) V_2 + \dots + C_{0m} \left(\frac{\sigma_m}{\sigma_0} \right) V_m + C_{0u} \left(\frac{\sigma_u}{\sigma_0} \right) V_u \quad (2)$$

Apabila $P_{0i} = C_{0i} \left(\frac{\sigma_i}{\sigma_0} \right)$ maka

$$V_0 = P_{01}V_1 + P_{02}V_2 + \dots + P_{0m}V_m + P_{0u}V_u \quad (3)$$

Dalam bentuk baku, semua koefisien korelasi direduksi dari hasil kali momen yaitu :

$$r_{0i} = \frac{\sum_{i=1}^m V_0 V_i}{n}$$

$$= P_{01}r_{1i} + P_{02}r_{2i} + \dots + P_{0m}r_{mi} + P_{0u}r_{ui} \tag{4}$$

$$= \sum_{i=1}^m \sum_{j=1}^m P_{0j}r_{ji}$$

dari persamaan (4), bila $i = 0$ maka

$$r_{00} = P_{01}r_{10} + P_{02}r_{20} + \dots + P_{0m}r_{m0} + P_{0u}r_{u0} \tag{5}$$

Jika $P_{0u} = r_{u0}$ dan karena $r_{00} = 1$ maka persamaan (5) menjadi

$$P_{0u} = \sqrt{1 - \sum_{j=1}^m P_{0j}r_{j0}}, \text{ yang merupakan rumus untuk galat untuk setiap variabel endogenus.}$$

Apabila variabel bebas yang dilibatkan dalam pembahasan hanya satu buah, sedangkan variabel bebas dan variabel sisa tak berkorelasi maka $r_{0i} = p_{0i}$.

Apabila koefisien jalur telah diperoleh, maka beberapa informasi akan diperoleh berdasarkan metode analisis jalur [8]. Informasi itu antara lain :

1. Pengaruh langsung (*Direct Effect / DE*).
2. Pengaruh tak langsung (*Indirect Effect / IE*).
3. Pengaruh galat (error) atau sisaan (residual).

Pengujian kecocokan model (*model fit*) dapat digunakan statistik *Chi-Square* yang dianjurkan oleh Specht (1975) dan Pedhazur (1982) [4]. Pengujian model juga dapat dilakukan dengan menggunakan indeks kesesuaian model yaitu RMSEA (*The Root Mean Square Error of Approximation*), GFI (*Goodness of Fit Index*), AGFI (*Adjusted Goodness of Fit Index*), CMIN (*The Minimum Sample Discrepancy Function*) / DF (*Degree of Freedom*), TLI (*Tucker Lewis Index*), dan CFI (*Comparative Fit Index*) [7]. Indeks-indeks yang dapat digunakan untuk menguji kelayakan sebuah model diringkas dalam Tabel berikut ini :

Tabel 1. *Goodness of Fit Indices*

| <i>Goodness of Fit</i> | <i>Cut of Value</i> |
|---------------------------------|---------------------|
| <i>Chi-Square</i> | Diharapkan kecil |
| <i>Significance Probability</i> | $\geq 0,05$ |
| RMSEA | $\leq 0,08$ |
| GFI | $\geq 0,90$ |
| AGFI | $\geq 0,90$ |
| CMIN/DF | $\leq 2,00$ |
| TLI | $\geq 0,95$ |
| CFI | $\geq 0,95$ |

METODE PENELITIAN

Penelitian ini merupakan penelitian terapan, yaitu penelitian yang bertujuan untuk memperoleh penemuan-penemuan yang berkenaan dengan aplikasi atau penerapan teori-teori tertentu [9]. Populasi dalam penelitian ini adalah 46 petani untuk sistem OT dan 72 petani untuk sistem TOT. Sampel ditentukan dengan menggunakan metode *simple random sampling* dengan cara estimasi proporsi sehingga diperoleh sampel untuk petani yang menggunakan sistem OT sebanyak 41 petani dan 61 petani untuk sistem TOT [5].

Data yang digunakan dalam penelitian ini adalah data sekunder yaitu data faktor-faktor yang mempengaruhi usahatani kacang tanah dengan sistem OT dan TOT di Desa Retak Mudik Kecamatan Perwakilan Pondok Suguh [5].

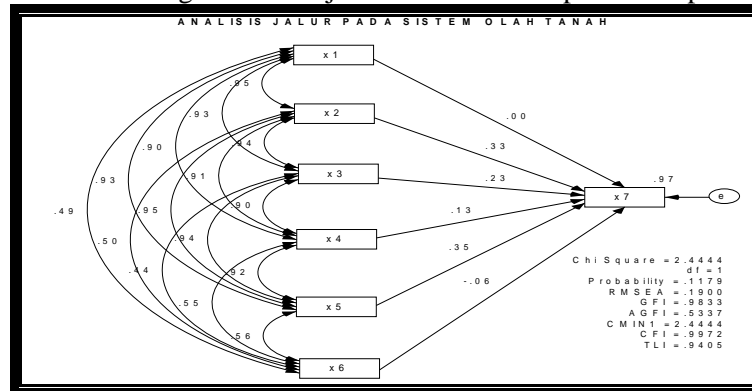
Variabel-variabel bebas yang digunakan dalam penelitian ini adalah : X_1 = Luas lahan (Ha), X_2 = Jumlah Tenaga kerja (HKSP), X_3 = Jumlah Pupuk TSP (Kg), X_4 = Jumlah Pupuk KCl (Kg), X_5 = Jumlah Benih (Kg), dan X_6 = Jumlah Pestisida Setara Toupan (L). Sedangkan variabel tak bebas dalam penelitian ini adalah X_7 = Jumlah Produksi. Penelitian ini dianalisis dengan langkah-langkah sebagai berikut :

1. Merancang model berdasarkan konsep dan teori.
2. Pemeriksaan terhadap asumsi yang melandasi analisis jalur.
3. Pendugaan parameter untuk kedua sistem pengolahan tanah.
4. Membandingkan pengaruh langsung dan pengaruh tidak langsung untuk kedua sistem pengolahan tanah.
5. Menentukan sistem pengolahan tanah yang terbaik untuk usahatani kacang tanah.

PEMBAHASAN

Analisis Jalur pada Sistem Olah Tanah

Besarnya koefisien jalur diperlihatkan oleh hasil output diagram jalur dengan menggunakan software AMOS 6. Harga koefisien jalur keseluruhan dapat dilihat pada Gambar 1.



Gambar 2 Analisis Jalur untuk Sistem Olah Tanah

Analisis atas Kesesuaian Model

Dari Gambar 2, dapat dilihat nilai signifikansi Chi-Square sebesar 2.4444. Dari tabel Chi-Square dengan $df = 1$ dan taraf signifikan sebesar $\alpha = 0.05$ didapat $\chi^2 = 3.84$. Karena $\chi^2_{hitung} = 2.4444 < \chi^2_{tabel} = 3.84$, menunjukkan bahwa model dapat diterima dengan baik. Dari Gambar 2 juga diperoleh nilai *probability* sebesar $0.1179 > 0.05$, nilai GFI sebesar $0.9833 > 0.90$ dan nilai CFI sebesar $0.9972 > 0.95$. Nilai-nilai ini menunjukkan bahwa model yang diuji memiliki *Goodness of Fit* yang baik sehingga model diatas dapat diterima.

Analisis atas Signifikansi Koefisien Jalur (*Path Coefficient*)

Analisis atas signifikansi koefisien jalur (*path coefficient*) dalam program AMOS 6 dapat dilihat melalui signifikansi besaran *regression weight* dari model seperti yang disajikan dalam Tabel 2.

Tabel 2. *Regression Weights* untuk Sistem Olah Tanah

| | Estimate | S.E. | C.R. | P | Label |
|------------|-----------|----------|----------|---------|-------|
| x7 <--- x1 | 1.00000 | | | | |
| x7 <--- x2 | 7.04524 | 1.98734 | 3.54506 | *** | par_1 |
| x7 <--- x3 | 5.39492 | 2.05298 | 2.62785 | 0.00859 | par_2 |
| x7 <--- x4 | 3.00060 | 1.73015 | 1.73430 | 0.08286 | par_3 |
| x7 <--- x5 | 9.23231 | 2.54984 | 3.62074 | *** | par_4 |
| x7 <--- x6 | -33.41925 | 17.54794 | -1.90445 | 0.05685 | par_5 |

Nilai C.R. (critical ratio) identik dengan t-hitung dalam analisis regresi. Oleh karena itu, nilai C.R. (critical ratio) yang lebih besar dari 2.0 menunjukkan adanya causal relationship yang signifikan.

Berdasarkan Tabel 2 di atas, variabel Luas Lahan (X_1), Jumlah Pupuk KCl (X_4), dan Jumlah Pestisida setara Toupan (X_6) mempunyai nilai C.R lebih kecil dari 2.0. Berarti variabel-variabel ini mempunyai pengaruh yang tidak signifikan terhadap Jumlah Produksi. Adapun variabel yang berpengaruh secara signifikan terhadap Jumlah Produksi, yaitu Jumlah Tenaga Kerja (X_2), Jumlah Pupuk TSP (X_3) dan Jumlah Benih (X_5).

Analisis atas Efek Total (Total Effect), Efek Langsung (Direct Effect), dan Efek Tidak Langsung (Indirect Effect)

Efek langsung, efek tidak langsung, dan efek total diperlihatkan dalam Tabel 3.

Tabel 3. Efek Langsung, Efek Tidak Langsung, dan Efek Total yang Dibakukan pada Sistem Olah Tanah

| | | x6 | x5 | x4 | x3 | x2 | x1 |
|---------------------|----|----------|---------|---------|---------|---------|---------|
| Efek Langsung | x7 | -0.06192 | 0.35478 | 0.12520 | 0.23372 | 0.32876 | 0.00064 |
| Efek Tidak langsung | x7 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 |
| Efek Total | x7 | -0.06192 | 0.35478 | 0.12520 | 0.23372 | 0.32876 | 0.00064 |

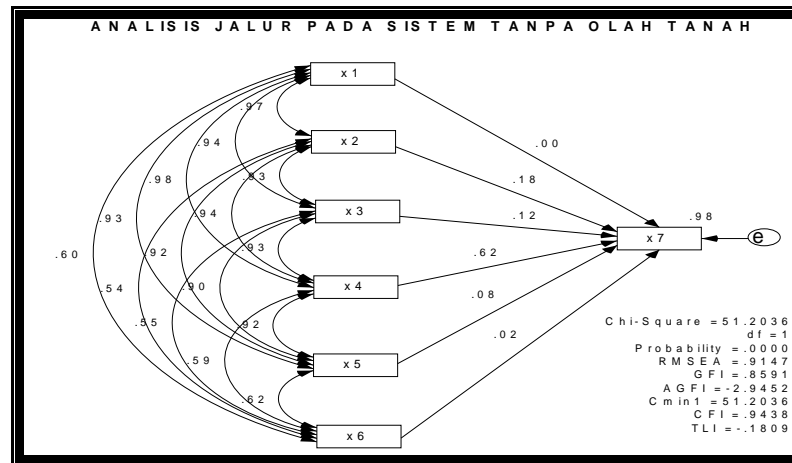
Berdasarkan Tabel 3, efek langsung terbesar diberikan oleh variabel Jumlah Benih (X_5) yaitu sebesar 0.35478. Sedangkan efek langsung terkecil diberikan oleh variabel Luas Lahan (X_1) yaitu sebesar 0.00064.

Berdasarkan Tabel 3, nilai dari masing-masing variabel yang mempengaruhi Jumlah Produksi sangat kecil (< 0.001). Ini menunjukkan bahwa tidak terdapat efek tidak langsung secara signifikan dari masing-masing variabel yang mempengaruhi Jumlah Produksi.

Berdasarkan Tabel 3, karena dalam model ini tidak terdapat efek tidak langsung dari masing-masing variabel yang mempengaruhi Jumlah Produksi, maka besar efek total sama dengan besar efek langsung dari masing-masing variabel yang mempengaruhi Jumlah Produksi.

Analisis Jalur pada Sistem Tanpa Olah Tanah (TOT)

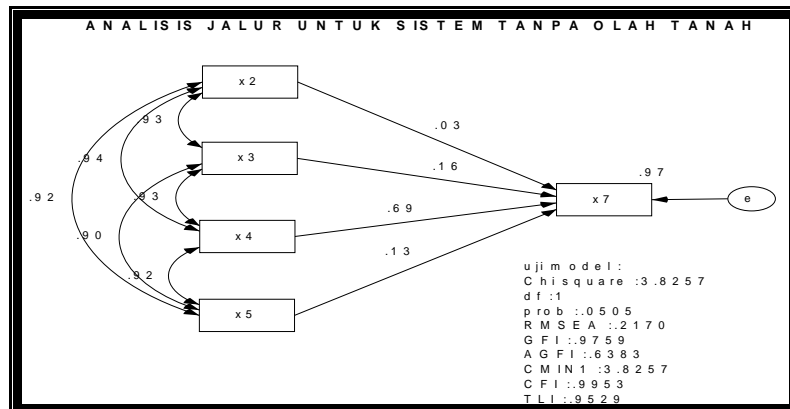
Besarnya koefisien jalur diperlihatkan oleh hasil output diagram jalur dengan menggunakan software AMOS 6. Harga koefisien jalur keseluruhan dapat dilihat pada Gambar 3.



Gambar 3 Analisis Jalur Untuk Sistem Tanpa Olah Tanah

Dari Gambar 3, dapat dilihat nilai signifikansi Chi-Square sebesar 51.2036. Karena $\chi^2_{hitung} = 51.2036 > \chi^2_{tabel} = 3.84$, menunjukkan bahwa model yang diperoleh kurang baik. Nilai - nilai indeks kesesuaian menunjukkan bahwa model yang diperoleh kurang baik.

Karena model yang diperoleh dari analisis ini kurang baik maka perlu dilakukan modifikasi model yaitu dengan menghilangkan jalur yang tidak signifikan ($< 0,05$). Dalam penelitian ini jalur yang tidak signifikan adalah jalur X_1 ke X_7 dan X_6 ke X_7 . Koefisien jalur ini dihilangkan dari model sehingga dihasilkan model sebagai berikut :



Gambar 4 Analisis Jalur untuk Sistem Tanpa Olah Tanah setelah Modifikasi Model

Terlihat bahwa nilai *Chi-Square* sebesar $3.8257 < \chi^2_{tabel} = 3.84$, menunjukkan bahwa model dapat diterima dengan baik. Dari Gambar 4 juga diperoleh nilai *probability* sebesar $0.0505 > 0.05$, nilai *GFI* sebesar $0.9759 > 0.90$, nilai *CFI* sebesar $0.9953 > 0.95$ dan nilai *TLI* sebesar 0.9529 . Nilai - nilai ini menunjukkan bahwa model yang diuji memiliki *Goodnes of Fit* yang baik sehingga model diatas dapat diterima.

Analisis atas Signifikansi Koefisien Jalur (*path coefficient*)

Analisis atas signifikansi koefisien jalur (*path coefficient*) dapat dilihat melalui signifikansi besaran *regression weight* dari model seperti yang disajikan dalam Tabel 6 berikut ini :

Tabel 6. Regression Weights untuk Sistem Tanpa Olah Tanah

| | Estimate | S.E. | C.R. | P | Label |
|------------|----------|---------|---------|---------|-------|
| x7 <--- x3 | 3.13248 | 1.22346 | 2.56035 | 0.01046 | par_7 |
| x7 <--- x5 | 2.63151 | 1.14498 | 2.29831 | 0.02154 | par_8 |
| x7 <--- x2 | 1.00000 | | | | |
| x7 <--- x4 | 15.33533 | 1.54108 | 9.95105 | *** | par_9 |

Berdasarkan Tabel 6 di atas, hasil pengujian ini menunjukkan bahwa semua jalur yang dianalisis setelah dilakukan modifikasi model mempunyai nilai C.R lebih besar dari 2.0. Berarti variabel Jumlah Tenaga Kerja (X_2), Jumlah Pupuk TSP (X_3), Jumlah Pupuk KCl (X_4) dan Jumlah Benih (X_5) berpengaruh secara signifikan terhadap Jumlah Produksi.

Analisis atas Efek Total (Total Effect), Efek Langsung (Direct Effect), dan Efek Tidak Langsung (Indirect Effect)

Efek langsung, efek tidak langsung, dan efek total dinyatakan dalam Tabel 7.

Tabel 7. Efek Langsung, Efek Tidak Langsung, dan Efek Total yang Dibakukan pada Sistem Tanpa Olah Tanah

| | | x5 | x4 | x3 | x2 |
|---------------------|----|---------|---------|---------|---------|
| Efek Langsung | x7 | 0.12948 | 0.68785 | 0.15503 | 0.03253 |
| Efek Tidak langsung | x7 | 0.00000 | 0.00000 | 0.00000 | 0.00000 |
| Efek Total | x7 | 0.12948 | 0.68785 | 0.15503 | 0.03253 |

Berdasarkan Tabel 7, efek langsung terbesar diberikan oleh variabel Jumlah Pupuk KCl (X_4) yaitu sebesar 0.68785. Sedangkan efek langsung terkecil diberikan oleh variabel Jumlah Tenaga Kerja (X_2) yaitu sebesar 0.03253.

Berdasarkan Tabel 7, nilai dari masing-masing variabel yang mempengaruhi Jumlah Produksi sangat kecil (< 0.001). Ini menunjukkan bahwa tidak terdapat efek tidak langsung secara signifikan dari masing-masing variabel yang mempengaruhi Jumlah Produksi.

Berdasarkan Tabel 7, karena dalam model ini tidak terdapat efek tidak langsung dari masing-masing variabel yang mempengaruhi Jumlah Produksi, maka besar efek total sama dengan besar efek langsung dari masing-masing variabel yang mempengaruhi Jumlah Produksi.

Perbandingan antara Sistem Olah Tanah dan Tanpa Olah Tanah

Untuk membandingkan kedua sistem tanam pada usahatani kacang tanah ini, dapat digunakan nilai koefisien jalur yang telah diperoleh dari analisis jalur pada kedua sistem tanam. Koefisien jalur untuk kedua sistem tanam tersebut disajikan dalam Tabel berikut :

Tabel 8. Koefisien Jalur untuk Kedua Sistem Tanam

| Variabel | Koefisien Jalur pada Sistem Olah Tanah | Koefisien Jalur pada Sistem Tanpa Olah Tanah |
|--------------------------------|--|--|
| Luas Lahan | 0.00064 | - |
| Jumlah Tenaga Kerja | 0.32876 | 0.03253 |
| Jumlah Pupuk TSP | 0.23372 | 0.15503 |
| Jumlah Pupuk KCl | 0.12520 | 0.68785 |
| Jumlah Benih | 0.35478 | 0.12948 |
| Jumlah Pestisida Setara Toupan | -0.06192 | - |

Berdasarkan koefisien jalur yang diperoleh, terlihat bahwa pada Sistem OT, variabel Luas Lahan, Jumlah Tenaga Kerja, Jumlah Pupuk TSP, Jumlah Benih dan Jumlah Pestisida Setara Toupan mempunyai nilai koefisien jalur yang lebih besar secara magnitud dibandingkan dengan nilai koefisien jalur pada Sistem TOT yaitu masing-masing sebesar 0.00064, 0.32876, 0.23372, 0.35478 dan -0.6192. Sehingga dapat dikatakan bahwa usahatani kacang tanah pada Sistem OT lebih baik dibandingkan pada sistem TOT di Desa Retak Mudik.

KESIMPULAN

1. Dari analisis jalur pada Sistem OT, model yang terbentuk cukup baik. Adapun variabel yang berpengaruh secara signifikan terhadap Jumlah Produksi, yaitu Jumlah Tenaga Kerja (X_2), Jumlah Pupuk TSP (X_3) dan Jumlah Benih (X_5). Efek langsung terbesar diberikan oleh variabel Jumlah Benih (X_5) yaitu sebesar 0.355. Sedangkan efek langsung terkecil diberikan oleh variabel Luas Lahan (X_1) yaitu sebesar 0.001.
2. Dari analisis jalur pada Sistem TOT, model yang terbentuk dari penelitian ini kurang baik. Tetapi setelah dilakukan modifikasi model diperoleh model yang cukup baik. Semua variabel hasil modifikasi model berpengaruh secara signifikan terhadap Jumlah Produksi. Efek langsung terbesar diberikan oleh variabel Jumlah Pupuk KCl (X_4) yaitu sebesar 0.68785. Sedangkan efek langsung terkecil diberikan oleh variabel Jumlah Tenaga Kerja (X_2) yaitu sebesar 0.03253.
3. Berdasarkan koefisien jalur yang diperoleh, dapat dikatakan bahwa usahatani kacang tanah pada Sistem Olah Tanah lebih baik dibandingkan pada Sistem Tanpa Olah Tanah di Desa Retak Mudik.

DAFTAR PUSTAKA

- [1] Anonim. 1999. *Path Analysis*.
www.luna.cas.usf.edu/~mbrannic/files/regression/Pathan.html. 13 Mei 2006.
- [2] Anonim. 2001. *Path Analysis*.
www2.chass.ncsu.edu/garson/pa765/path.html. 12 Mei 2006.
- [3] Anonim. 2005. *Pengembangan Analisis Multivariat dengan SPSS 12*. Jakarta : Salemba Infotek.
- [4] Bachrudin, A. dan Harapan L. Tobing. 2003. *Analisis Data untuk Penelitian Survei dengan Menggunakan Lisrel 8, Dilengkapi Contoh Kasus*. Bandung : Universitas Padjajaran.
- [5] Busrah, E. 2004. *Analisis Produksi dan Pendapatan Usahatani Kacang Tanah dengan Sistem Olah Tanah dan Tanpa Olah Tanah serta Pemasarannya (Studi Kasus : Desa Retak Mudik Kecamatan Pondok Suguh Bengkulu Utara)*. Skripsi FP. Tidak dipublikasikan.
- [6] Carey, G. 1998. *Multiple Regression and Path Analysis*.
www.exeter.ac.uk/~SEGLEa/multvar2/pathanal.html. 12 Mei 2006
- [7] Ferdinand, A. 2002. *Structural Equation Modeling dalam Penelitian Manajemen*. Semarang : BP UNDIP
- [8] Gaspersz, V. 1992. *Teknik Analisis dalam Penelitian Percobaan*. Bandung: Tarsito.
- [9] Mardalis. 1989. *Metode Penelitian Suatu Pendekatan Proposal*. Jakarta : Bumi Aksara.
- [10] Sudjana. 2002. *Teknik Analisis Regresi dan Korelasi Bagi Para Peneliti*. Bandung: Tarsito.
- [11] Wibowo, A. 2005. *Pengantar Analisis Jalur (Path Analysis)*. Surabaya : Lembaga Penelitian Universitas Airlangga.

Pengklasteran Provinsi Berdasarkan Persentase Penduduk yang Mempunyai Keluhan Kesehatan dalam Sebulan pada Tahun 2003

Lisa Noviyanti¹⁾, Sigit Nugroho²⁾, Fachri Faisal²⁾

¹⁾Alumni Jurusan Matematika FMIPA Universitas Bengkulu

²⁾Staf Pengajar Jurusan Matematika FMIPA Universitas Bengkulu

ABSTRAK

Strategi kebijakan pembangunan kesehatan baru yang dikenal dengan visi pembangunan kesehatan yang secara singkat dinyatakan sebagai “Indonesia Sehat 2010” bertujuan untuk mengoptimalkan derajat kesehatan. Derajat kesehatan yang optimal merupakan salah satu faktor untuk mensukseskan pembangunan nasional. Derajat kesehatan adalah tingkat kesehatan perorangan, kelompok atau masyarakat yang diukur dengan angka kesakitan, status gizi, angka kematian dan harapan hidup. Penelitian deskriptif eksploratif ini bertujuan untuk mengklasterkan provinsi di Indonesia berdasarkan persentase penduduk yang mempunyai keluhan kesehatan dalam sebulan pada tahun 2003. Pengklasteran dilakukan dengan metode *k-means clustering* dan hirarki aglomeratif yaitu: *single linkage*, *complete linkage*, metode ward, median, dan sentroid. Pengklasteran ini menghasilkan 5 klaster yang ideal..

Key words : *cluster analysis, dendogram, aglomerative, k-means clustering*

PENDAHULUAN

Pemerintah Republik Indonesia telah menyusun strategi kebijakan pembangunan kesehatan baru. Kebijakan ini didasarkan pada gerakan pembangunan berwawasan kesehatan sebagai Strategi Nasional menuju Indonesia Sehat 2010 yang dicanangkan oleh Presiden Republik Indonesia pada pembukaan Rakernas Departemen Kesehatan RI pada tanggal 1 Maret 1999.

Dengan rumusan ini diharapkan pada tahun 2010 bangsa Indonesia sudah hidup dalam lingkungan yang sehat, berperilaku hidup bersih dan sehat, dapat memilih, menjangkau dan memanfaatkan pelayanan kesehatan yang bermutu, merata dan berkeadilan sehingga memiliki derajat kesehatan yang optimal (Anonim, 2004).

Meningkatnya jumlah penduduk berarti upaya untuk meningkatkan derajat kesehatan masyarakat perlu diperhatikan. Hasil penelitian United Development Project (UNDP) menyatakan bahwa *Human Development Indeks* (indeks pembangunan manusia) Indonesia berada pada urutan ke 117 dari 175 negara. Oleh karena itu, masalah kesehatan tidak tepat lagi bila dipandang sebagai suatu konsumsi tapi sebagai investasi dalam upaya mencapai kesejahteraan masyarakat (Argadiredha, 2003).

Berbagai upaya untuk meningkatkan derajat kesehatan masyarakat telah dilakukan. Membuat dan melaksanakan program pembangunan di bidang kesehatan agar merata bagi segenap masyarakat Indonesia. Namun masih disadari bahwa masih terjadi ketimpangan sehingga derajat kesehatan masyarakat masih rendah dan sangat tertinggal jika dibandingkan dengan negara-negara tetangga (Anonim, 2002a).

Salah satu indikator yang digunakan untuk menentukan derajat kesehatan tersebut adalah angka kesakitan (Anonim, 2003). **Derajat Kesehatan** adalah tingkat kesehatan perorangan, kelompok atau masyarakat yang diukur dengan angka kesakitan, umur harapan hidup, status gizi dan angka kematian. Banyaknya orang yang sakit menggambarkan kondisi kesehatan suatu bangsa

(Anonim, 1991). Angka kesakitan ini diharapkan dapat menurun sesuai dengan sasaran program pokok upaya kesehatan baik untuk penyakit menular atau penyakit tidak menular.

Analisis klaster merupakan suatu teknik yang dapat digunakan untuk mengklasifikasikan obyek (provinsi-provinsi di Indonesia) ke dalam kelompok yang relatif homogen yang disebut klaster. Pengklasteran berdasarkan persentase penduduk yang mempunyai keluhan kesehatan dapat memberikan informasi berupa profil sehingga dapat membantu mencari faktor penyebab timbulnya masalah kesehatan. Dengan demikian hasil pengklasteran dapat membantu pemerintah dalam mensukseskan program Indonesia Sehat 2010.

Berdasarkan hal tersebut maka penulis tertarik untuk mengklasterkan provinsi di Indonesia berdasarkan persentase penduduk yang mempunyai keluhan kesehatan dalam sebulan pada tahun 2003.

Bagaimana membuat pengklasteran provinsi di Indonesia berdasarkan persentase penduduk yang mempunyai keluhan kesehatan dalam sebulan pada tahun 2003 yang dapat memberikan profil yang menjelaskan karakteristik masing-masing provinsi di Indonesia.

Tujuan penelitian ini adalah mengklasterkan provinsi-provinsi di Indonesia berdasarkan persentase penduduk yang mempunyai keluhan kesehatan selama sebulan pada tahun 2003 sehingga provinsi-provinsi yang mempunyai kemiripan akan berada dalam satu klaster.

Penelitian ini dapat memberikan informasi berupa gambaran tentang kesehatan masyarakat dan profil yang menjelaskan karakteristik setiap klaster. Profil dapat digunakan sebagai sarana diagnosis dalam mencari sebab masalah kesehatan sehingga memudahkan pencarian pemecahan masalah. Dengan demikian hasil pengklasteran dapat membantu mengoptimalkan derajat kesehatan bangsa Indonesia dan mensukseskan program Indonesia Sehat 2010.

TINJAUAN PUSTAKA

Indonesia Sehat 2010

Tujuan pembangunan kesehatan (Undang-Undang Nomor 23 tahun 1992) adalah untuk meningkatkan kesadaran, kemauan dan kemampuan bagi tiap warga negara Indonesia agar berada pada kondisi kesehatan yang optimal. Hal senada juga tertuang dalam visi pembangunan kesehatan dan Garis-Garis Besar Haluan Negara (GBHN) 1999-2004 yang mengamanatkan bahwa pembangunan kesehatan diarahkan untuk dapat meningkatkan derajat kesehatan, meningkatkan mutu dan kemudahan pelayanan kesehatan agar terjangkau oleh seluruh lapisan masyarakat. Termasuk juga, meningkatkan keadaan gizi, membudayakan pola hidup bersih dan sehat (PHBS), didukung oleh pembangunan pemukiman yang layak yang dikelola melalui manajemen yang efisien dalam pemenuhan kebutuhan aspirasi masyarakat sasaran (Anonim, 2002a).

Derajat kesehatan atau proporsi penduduk dalam satu wilayah yang mengalami sakit selama periode tertentu adalah tingkat kesehatan perorangan, kelompok atau masyarakat. Dalam hal ini, keluhan kesehatan selama sebulan yang lalu sebelum survei. Keluhan kesehatan digambarkan lewat angka kesakitan macam-macam penyakit (persentase penduduk yang mengalami kesakitan selama periode tertentu) masing-masing jenis keluhan penduduk. Keluhan kesehatan adalah keadaan seseorang yang mengalami gangguan kesehatan dan kejiwaan baik karena penyakit akut, penyakit kronis, kecelakaan, kriminal atau hal lain (Anonim, 2003a).

Analisis Klaster

Analisis klaster merupakan salah satu analisis multivariat yang termasuk dalam metode interdependensi yaitu variabel bebas x atau faktor penyebab tidak dibedakan dengan variabel terikat y atau respon (Supranto, 2004).

Istilah analisis klaster pertama kali digunakan oleh Tryon (1939) (Anonim, 2004b). Analisis klaster adalah suatu koleksi metode statistik yang mengidentifikasi kelompok sampel berdasarkan

karakteristik serupa (Sambamorthi, 2005). Analisis kluster mengelompokkan elemen mirip sebagai obyek penelitian yang mempunyai tingkat homogenitas yang tinggi antar obyek menjadi kluster yang berbeda dengan tingkat heterogenitas obyek yang tinggi antar kluster. Pengklasteran ini didasarkan pada gugus variabel yang diper timbangkan untuk diteliti.

Hasil pengklasteran diharapkan akan menyediakan beberapa pengertian yang mendalam untuk masing-masing provinsi di Indonesia. Pengklasteran mempunyai efek mengurangi banyak data dengan mengurangi banyaknya obyek (Anonim, 2005b). Analisis kluster harus memenuhi asumsi berikut (Santoso, 2002):

- a. Sampel yang diambil harus benar-benar bisa mewakili populasi
- b. Multikolinieritas yaitu korelasi antar obyek. Sebaiknya tidak ada, bila ada maka besar multikolinieritas tidaklah tinggi ($< 0,5$).

Sebelum melakukan analisis kluster, data yang digunakan juga perlu diperhatikan. Apakah terdapat perbedaan nilai yang besar antar variabel. Perbedaan data yang besar akan menyebabkan perhitungan jarak menjadi tidak valid sehingga data harus ditransformasi. Transformasi dapat dilakukan dengan uji *z-score*.

Untuk menentukan kedua obyek dikatakan mirip, perlu didefinisikan ukuran kemiripan antar dua obyek. Hal ini dilakukan untuk memperoleh matrik *proximity* yaitu matrik persegi dan simetri dengan jumlah obyek yang sama pada baris dan kolom. Matrik ini menunjukkan kemiripan atau ketakmiripan antar obyek (Anonim, 2003c).

Adapun metode yang dapat digunakan untuk mengukur kesamaan antar obyek yaitu (Santoso, 2004)

1. Mengukur jarak antar dua obyek

Metode ini berbentuk matrik simetri $n \times n$ yang berisi kemiripan atau ketakmiripan antar obyek sehingga jarak antar dua obyek bisa langsung diukur.

2. Mengukur korelasi antar sepasang obyek pada beberapa variabel

Pada metode ini datanya berbentuk matrik. Kesamaan antar obyek didapat dengan transformasi satu-satu sehingga indeks ketakmiripan bisa dikonversi menjadi indeks kemiripan. Salah satu yang jelas bisa menjadi ukuran ketakmiripan adalah fungsi jarak antara obyek a dan b di tulis $d(a,b)$. Fungsi ini harus memenuhi (Anonim (2003b), Portier (2003))

$$d(a,b) \geq 0$$

$$d(a,a) = 0$$

$$d(a,b) = d(b,a)$$

(a,b) meningkat seiring tidak miripnya a dan b

$$d(a,c) \leq d(a,b) + d(b,c)$$

Pengukuran jarak ada bermacam-macam namun yang paling sering digunakan adalah jarak *euclid* yaitu jarak geometris di ruang multidimensional (Anonim, 2004b). Beberapa cara dalam mengukur jarak yaitu

- a. Menggunakan jarak *euclid* yaitu akar jumlah kuadrat perbedaan nilai untuk tiap variabel (Wichern, 2002).

Jika $\mathbf{x}' = (x_1, x_2, \dots, x_p)$ dan $\mathbf{y}' = (y_1, y_2, \dots, y_p)$ maka

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_p - y_p)^2}$$

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{(\mathbf{x} - \mathbf{y})'(\mathbf{x} - \mathbf{y})} = \sqrt{\sum_{i=1}^p (x_i - y_i)^2}$$

b. Menggunakan jarak kuadrat *euclid*

c. *The City Block or Manhattan Distance* antara dua obyek merupakan jumlah nilai perbedaan mutlak untuk tiap variabel. Jarak ini juga disebut jarak Minkowski (Dillon & Goldstein, 1984).

Jika $\mathbf{x}' = (x_1, x_2, \dots, x_p)$; p adalah variabel

maka $\mathbf{x}'_i = (x_{i1}, x_{i2}, \dots, x_{ip})$ adalah kumpulan variabel pada obyek ke i

$$d_{ij} = \left[\sum_{k=1}^p |x_{ik} - y_{jk}|^r \right]^{\frac{1}{r}}$$

d_{ij} adalah jarak antara obyek ke i dan obyek ke j .

d. *The Chebyshev Distance* antar dua obyek ialah nilai perbedaan mutlak yang maksimum pada tiap variabel (Supranto, 2004).

3. Mengukur asosiasi antar obyek

Pada metode ini, data berbentuk non matrik (nominal atau ordinal). Jika akan dibandingkan obyek r dan obyek s maka indeks kesamaan ditunjukkan pada tabel 1 (Portier, 2003).

Tabel 1 Kesamaan Antara Dua Obyek

| Obyek r | obyek s | | |
|---------|---------|-------|-------------|
| | 1 | 0 | Jml |
| 1 | a | b | $a+b$ |
| 0 | c | d | $c+d$ |
| Jml | $a+c$ | $b+d$ | $a+b+c+d=p$ |

Asumsikan 1 jika karakteristik ada dan 0 jika karakteristik tidak ada. Koefisien kesamaan didefinisikan pada tabel 2.

Tabel 2 Koefisien Kesamaan Antara Dua Obyek

| Nama Koefisien | Persamaan |
|----------------|--|
| Jaccard | $S_{rs} = \frac{a}{(a+b+c)}$ |
| Dice | $S_{rs} = \frac{2a}{(2a+b+c)}$ |
| Ochiai | $S_{rs} = \frac{a}{[(a+b)(a+c)]^{\frac{1}{2}}}$ |
| Russell-Rao | $S_{rs} = \frac{a}{p}$ |
| Yule | $S_{rs} = \frac{(ad-bc)}{ad+bc}$ |
| Phi | $S_{rs} = \frac{(adbc)}{[(a+b)(a+c)(b+d)(c+d)]^{\frac{1}{2}}}$ |

Konversi ketaksamaan untuk analisa $d_{rs} = 1 - S_{rs}$

d_{rs} = Koefisien ketaksamaan antara obyek r dan obyek s

S_{rs} = Koefisien kesamaan antara obyek r dan obyek s

Metode Pengklasteran

Metode Hirarki

Metode hirarki bisa aglomeratif (*agglomeration*) atau devisif (*devisife*) (Anonim, 2005b). Metode Aglomeratif terdiri dari :

1. Linkage Method (Metode Linkage)

Metode ini dibagi lagi menjadi tiga metode yaitu (Wichern, 2002):

i. Single linkage

Single linkage dan *complete linkage* kali pertama didiskusikan oleh Sneath (1957) lalu dibahas lagi oleh Sokal Sneath Classic (1963). Kemudian ke dua metode ini juga didiskusikan oleh Johnson (1967), (Dillon & Goldstein, 1984). Jika jarak antara kluster B_r dan B_s adalah $h(B_r, B_s)$ didefinisikan sebagai berikut:

$$h(B_r, B_s) = \min \{ d(x_i, x_j); x_i \text{ anggota } B_r \text{ dan } x_j \text{ anggota } B_s \}$$

Kluster B_r dan B_s akan digabung jika $h(B_r, B_s)$ adalah jarak yang terkecil sehingga metode ini juga disebut aturan tetangga dekat (Dillon & Goldstein, 1984).

ii. Complete linkage

Metode ini hampir sama dengan *single linkage* hanya saja pada metode ini menggunakan jarak yang paling jauh antara dua kluster yang berbeda B_r dan B_s yang didefinisikan sebagai berikut:

$$h(B_r, B_s) = \max \{ d(x_i, x_j); x_i \text{ anggota } B_r \text{ dan } x_j \text{ anggota } B_s \}$$

iii. Average linkage

Metode ini menggunakan rata-rata jarak antara semua pasangan obyek sebagai jarak antara dua kluster. Metode ini jarang digunakan jika dibandingkan *single linkage* dan *complete linkage* karena metode ini membutuhkan informasi pada semua pasangan jarak. Namun ke tiga metode *linkage* ini seringkali memberikan hasil yang hampir sama (Anonim, 2003c).

Jarak antara dua kluster B_r dan B_s didefinisikan sebagai berikut:

$$h(B_r, B_s) = \frac{1}{n_1 n_2} \sum_{x_i \in B_r} \sum_{x_j \in B_s} d(x_i, x_j)$$

2. Ward method (Metode Ward)

Ward (1963) mengusulkan penggunaan metode yang didasarkan pada hasil informasi yang minimum dari kenaikan pada jumlah kuadrat deviasi rata-rata kluster. Proses berhenti pada kenaikan yang menyebabkan *error sum of squares* (ESS) dari gabungan tiap kluster yang mungkin. Nilai ESS digunakan sebagai fungsi obyektif dan didefinisikan sebagai berikut (Dillon & Goldstein, 1984):

$$ESS = \sum_{j=1}^k \left(\sum_{i=1}^{n_j} x_{ij}^2 - \frac{1}{n_j} \left(\sum_{i=1}^{n_j} x_{ij} \right)^2 \right)$$

x_{ij} : Nilai obyek ke i pada kluster ke j

k : Jumlah kluster tiap *stage*

n_j : Jumlah obyek ke i pada kluster ke j

Metode ini juga dikenal dengan metode varian minimum dan harus menggunakan jarak kuadrat *euclid* namun sulit untuk menggunakannya tanpa bantuan komputer (Anonim, 2005c).

3. Centroid Method (Metode Centroid)

Sokal dan Michener (1958), King (1996, 1997) dalam Seber (1984) menyatakan bahwa jarak antara dua kluster didefinisikan sebagai jarak *euclid* antar kedua rata-rata (*centroid*) kluster. Jika \bar{x}_r dan \bar{x}_s

adalah vektor rata-rata (*centroid*) kluster B_r dan B_s , maka jarak antar dua kluster didefinisikan sebagai $h(B_r, B_s) = d(\bar{x}_r, \bar{x}_s)$. *Centroid* kluster baru yang terbentuk didapat dengan rumus

$$\frac{n_r \bar{x}_r + n_s \bar{x}_s}{n_r + n_s}$$

n_r dan n_s adalah banyaknya anggota kluster B_r dan B_s .

4. Median Method (Metode Median)

Terkadang terdapat ukuran kluster B_s jauh lebih kecil dari pada kluster B_r yaitu $n_s \ll n_r$ bila kedua kluster digabungkan maka *centroid* dari kluster baru tidak akan jauh berbeda dengan \bar{x}_r . Untuk menghindari kontribusi B_s terhadap pembentukan jarak yang baru tidak terlalu besar, Gower menyarankan penggunaan *median* antara kluster yang digabungkan sebagai titik untuk menghitung jarak yang baru.

Jika kluster B_r dan B_s digabung maka akan diperoleh median baru yang didefinisikan sebagai berikut (Anonim, 2003b):

$$m_{baru} = \frac{m_r + m_s}{2}$$

Median ini dihitung sebagai titik tengah pada garis yang menghubungkan median lama dan median baru. Dengan demikian jarak antar kluster didefinisikan sebagai jarak antar median.

Metode berhirarki devisif merupakan kebalikan metode aglomeratif. Metode pengklasteran ini dimulai dengan menganggap bahwa hanya ada satu kluster yang memuat semua obyek.

William dan Lance dalam Seber (1984) menyatakan metode devisif lebih menguntungkan dibandingkan aglomeratif. Namun metode ini jarang digunakan (Anonim, 2005c) dan tidak semua *software* menyediakan fasilitas metode ini. Oleh karena itulah, metode berhirarki devisif tidak penulis bahas pada skripsi ini.

Metode Non Hirarki (Pengklasteran *K-means*)

Berbeda dengan metode hirarki, metode ini justru dimulai dengan menentukan terlebih dahulu jumlah kluster yang diinginkan dan *centroid* di tiap kluster. Pada beberapa *software*, *centroid* yang digunakan adalah k pengamatan pertama namun ada juga *software* yang menentukan *centroid* secara acak. Pengklasteran *K-means* pertama kali dipopulerkan oleh Hartigan pada tahun 1975 (Anonim, 2004b).

Venable & Ripley (1995), Anonim (2003b) dan Wichern (2002) menyatakan bahwa Mac Queen menyarankan penggunaan pengklasteran *K-means* untuk menjelaskan algoritma dalam penentuan suatu obyek ke dalam kluster tertentu berdasarkan rata-rata terdekat.

Asumsikan n adalah obyek dan p adalah variabel yang dinotasikan dengan $x(i, j)$ $i=1, 2, \dots, n$ dan $j=1, 2, \dots, p$ dan dengan menggunakan jarak *euclid* antar obyek. Jika $p(n, k)$ adalah partisi yang merupakan hasil pada tiap obyek dialokasikan untuk salah satu dari kluster ke $1, 2, \dots, k$. Rata-rata variabel ke j pada kluster ke l dinotasikan dengan $\bar{x}(i, j)$, dan jumlah obyek pada kluster ke l dinotasikan dengan $n(l)$. Maka jarak antara obyek ke i dan kluster ke l didefinisikan sebagai berikut:

$$d(i, l) = \left(\sum_{j=1}^p [x(i, j) - \bar{x}(i, j)]^2 \right)^{1/2}$$

dengan

$$E[p(n, k)] = \sum_{i=1}^n D[i, l(i)]^2$$

adalah *error* partisi. $l(i)$ adalah kluster yang memuat obyek ke i , $D[i,l(i)]$ adalah jarak *euclid* antara obyek ke i dan rata-rata kluster yang memuat obyek.

Output metode hirarki adalah semacam ringkasan yang digambarkan oleh dendogram. Dendogram merupakan diagram seperti pohon dua dimensi yang mengilustrasikan pemisahan atau penggabungan dengan tingkat yang berjenjang (Dillon & Goldstein, 1984).

Pada metode hirarki, dendogram dapat membantu peneliti untuk menentukan jumlah kluster yang ideal. Posisi garis pada skala menunjukkan jarak. Jarak pada tahapan awal mempunyai nilai yang hampir sama sehingga sukar untuk menentukan urutan beberapa kluster awal dibentuk. Namun, jelas sekali bahwa dua tahap terakhir mempunyai jarak yang besar untuk digabung. Sebenarnya tidak ada aturan baku dalam menentukan banyak kluster tergantung subjektivitas peneliti. Peneliti juga dapat menggunakan pertimbangan teoritis, konseptual dan praktis.

Interpretasi Profil dan Akses Validitas Kluster

Interpretasi profil kluster meliputi pengkajian mengenai *centroid* yaitu rata-rata nilai obyek yang terdapat dalam kluster pada tiap variabel. Nilai *centroid* memungkinkan untuk menguraikan setiap kluster dengan cara memberi suatu label atau nama. Label suatu kluster juga dapat didasarkan pada manfaat yang akan di cari.

Pengecekan mutu hasil pengklasteran dapat dilakukan dengan analisis kluster yang menggunakan ukuran jarak yang berbeda, menggunakan metode pengklasteran yang berbeda dan membandingkan hasilnya.

METODE PENELITIAN

Penelitian ini merupakan penelitian deskriptif yang bersifat eksploratif dengan variabel adalah macam-macam keluhan kesehatan yang dialami penduduk selama sebulan sebelum survei dilakukan pada tahun 2003. Adapun variabelnya yaitu

1. Panas
2. Sakit Kepala
3. Batuk
4. Pilek
5. Diare
6. Sesak Napas
7. Asma
8. Sakit Gigi
9. Kejang-Kejang
10. Lumpuh
11. Telinga Berair
12. Campak
13. Sakit Kuning
14. Pikun
15. Kecelakaan
16. Lainnya

Data diperoleh dari buku Statistik Potensi Desa 2003 terbitan Badan Pusat Statistik (BPS) tahun 2003. Data tersebut diperoleh BPS dari hasil pengolahan data di tingkat provinsi.

Data dianalisis melalui tahap sebagai berikut

1. Menilai perlunya melakukan transformasi data
2. Analisis kluster
 - a. Metode hirarki
 - b. Metode pengklasteran *K-means*
3. Interpretasi profil kluster
4. Akses validitas kluster

HASIL DAN PEMBAHASAN

Pendahuluan

Pada data tersebut tidak terlihat adanya variasi dalam satuan atau tidak terdapat perbedaan yang mencolok. Perbedaan data yang besar akan membuat perhitungan menjadi tidak valid (Santoso, 2004). Oleh karena itu, data tidak perlu ditransformasi.

Pengklasteran dilakukan dengan dua metode yaitu metode hirarki dan metode pengklasteran *K-means*. Metode – metode yang terdapat dalam metode hirarki digunakan sebagai bahan eksplorasi dan perbandingan bagi peneliti. Walaupun metode – metode hirarki tersebut mempunyai fungsi dan peran yang sama tetapi terkadang berbeda dalam hasil pengklasterannya.

Hasil Uji Asumsi

Penelitian ini menggunakan total sampling yaitu populasi. Dengan demikian asumsi bahwa sampel yang digunakan harus benar benar bisa mewakili populasi telah dipenuhi. Multikolinieritas terjadi bila terdapat dua atau lebih variabel yang saling berkaitan. Untuk mengetahui ada tidaknya multikolinieritas antar variabel dapat dilakukan dengan matrik korelasi (Lampiran 2).

Pada tabel matrik korelasi dapat dilihat bahwa masih terdapat nilai korelasi yang lebih besar dari 0,5. Variabel tersebut adalah batuk dan pilek (0,636), sakit kepala dan lumpuh (0,6155), sakit kepala dan diare (0,59766), panas dan kejang (0,55661).

Dengan menggunakan $\alpha=0,001$ pada ke empat pasang variabel tersebut diperoleh hasil bahwa H_0 diterima. Artinya, variabel-variabel tersebut *independent* atau tidak terdapat hubungan antar variabel. Dengan Demikian tidak terdapat multikolinieritas.

Metode Hirarki

Karena data berbentuk matrik dan jumlah obyek tidak sama dengan jumlah variabel maka metode yang digunakan adalah mengukur korelasi antar sepasang obyek pada beberapa variabel.

Nilai matrik *proximity* tergantung dari jarak yang digunakan. Metode ward harus menggunakan jarak kuadrat *euclid* begitu juga dengan metode median dan metode *centroid* (Anonim (2005c) dan output SPSS versi 11.0). Sedangkan metode *linkage* tidak ada sumber yang mengharuskan menggunakan jarak tertentu namun jarak yang paling sering digunakan adalah jarak *euclid*. Nilai matrik *proximity* terlampir untuk jarak *euclid* dan jarak kuadrat *euclid* (Lampiran 3 dan lampiran 4). Semakin kecil nilai pada tabel matrik *proximity* berarti semakin mirip kedua obyek dalam karakteristiknya (keluhan kesehatan masyarakat dalam sebulan pada tahun 2003). Sebaliknya, semakin besar nilai jarak pada tabel matrik *proximity* maka semakin berbeda kedua obyek tersebut.

Tabel 3 Aglomerasi Metode *Single linkage*

| Stage | Cluster Combined | | Coefficients | Stage Cluster First Appears | | Next Stage |
|-------|------------------|-----------|--------------|-----------------------------|-----------|------------|
| | Cluster 1 | Cluster 2 | | Cluster 1 | Cluster 2 | |
| 1 | 3 | 5 | 4,126 | 0 | 0 | 4 |
| 2 | 14 | 22 | 5,139 | 0 | 0 | 4 |
| 3 | 6 | 13 | 5,485 | 0 | 0 | 6 |
| 4 | 3 | 14 | 5,532 | 1 | 2 | 9 |
| 5 | 25 | 26 | 5,6 | 0 | 0 | 21 |
| 6 | 6 | 11 | 5,68 | 3 | 0 | 8 |
| 7 | 15 | 19 | 5,935 | 0 | 0 | 10 |
| 8 | 6 | 12 | 6,125 | 6 | 0 | 9 |
| 9 | 3 | 6 | 6,711 | 4 | 8 | 10 |
| 10 | 3 | 15 | 6,799 | 9 | 7 | 11 |
| 11 | 3 | 7 | 6,877 | 10 | 0 | 13 |
| 12 | 4 | 20 | 6,961 | 0 | 0 | 14 |
| 13 | 1 | 3 | 7,231 | 0 | 11 | 14 |
| 14 | 1 | 4 | 7,698 | 13 | 12 | 15 |
| 15 | 1 | 2 | 7,715 | 14 | 0 | 16 |
| 16 | 1 | 10 | 7,758 | 15 | 0 | 17 |
| 17 | 1 | 30 | 7,982 | 16 | 0 | 18 |
| 18 | 1 | 9 | 9,18 | 17 | 0 | 20 |
| 19 | 28 | 29 | 9,594 | 0 | 0 | 26 |
| 20 | 1 | 21 | 9,847 | 18 | 0 | 21 |
| 21 | 1 | 25 | 9,86 | 20 | 5 | 22 |
| 22 | 1 | 23 | 9,993 | 21 | 0 | 23 |
| 23 | 1 | 17 | 10,415 | 22 | 0 | 24 |
| 24 | 1 | 16 | 10,419 | 23 | 0 | 25 |
| 25 | 1 | 24 | 12,594 | 24 | 0 | 26 |
| 26 | 1 | 28 | 14,229 | 25 | 19 | 27 |
| 27 | 1 | 18 | 14,634 | 26 | 0 | 28 |
| 28 | 1 | 27 | 14,67 | 27 | 0 | 29 |
| 29 | 1 | 8 | 51,198 | 28 | 0 | 0 |

Tabel aglomerasi diatas menunjukkan cara penggabungan kluster tiap tahap pada metode *single linkage* dengan menggunakan jarak *euclid*. Tabel aglomerasi metode lain dapat dilihat pada lampiran 5.

Pada tahap pertama terbentuk satu kluster dengan anggota obyek 3 (Sumatra Barat (Sumbar)) dan obyek 5 (Jambi). Kolom koefisien menyatakan jarak antara Sumatra Barat dan Jambi seperti yang terlihat pada tabel matrik *proximity* pada lampiran 3. Nilai koefisien tersebut didapat dari jarak yang digunakan. Karena proses aglomerasi dimulai dengan dua obyek terdekat maka kedua obyek tersebut adalah yang terdekat dari 30 kombinasi jarak 30 provinsi. Se makin kecil nilai koefisien pada tabel aglomerasi berarti semakin mirip satu obyek dengan obyek yang lain, sebaliknya semakin besar nilai koefisien pada tabel aglomerasi berarti semakin tak mirip satu obyek dengan obyek yang lain.

Kolom *next stage* menyatakan tahapan suatu obyek digabung dengan kluster yang baru saja dibentuk. Pada tahap pertama terlihat angka 4. Hal ini berarti langkah pengklasteran berikutnya dengan melihat langkah 4. Demikian seterusnya sampai langkah terakhir yakni langkah ke 29.

Output metode kluster digambarkan seperti diagram pohon yang dikenal dengan nama dendogram. Dendogram merupakan visualisasi proses pengklasteran yang terjadi. Dendogram masing-masing metode hirarki terlampir (Lampiran 6).

Dendogram dibaca dari kiri ke kanan. Posisi garis pada skala menunjukkan jarak suatu kluster digabung. Skala yang digunakan bukanlah koefisien yang ada pada tabel aglomerasi namun telah dilakukan penskalaan ulang dengan batasan 0 sampai 25. Garis vertikal mewakili kluster yang digabung bersama. Dendogram berguna untuk menunjukkan anggota kluster yang ada jika akan ditentukan berapa kluster yang seharusnya dibentuk. Dari dendogram dapat dilihat bahwa dua tahap terakhir, jarak suatu kluster digabung angkanya besar. Informasi ini berguna untuk memutuskan banyaknya kluster.

Dari dendogram metode *single linkage* dengan jarak *euclid* dapat dilihat bahwa ada empat kluster awal yang terbentuk dengan jarak yang paling dekat yaitu (Sumbar, Jambi, Jawa Timur (Jatim) dan Kalimantan Timur (Kaltim)), (Sumatra Selatan (Sumsel), Yogya dan Jawa Barat (Jabar)), (Banten dan Kalimantan Barat (Kalbar)) dan (Sulawesi Selatan (Sulsel) dan Sulawesi Tenggara (Sultra)). Kluster terakhir yang terbentuk yaitu kluster yang paling jauh jaraknya atau provinsi yang paling tidak mirip dibandingkan provinsi lain. Kluster ini adalah kluster provinsi Lampung kemudian diikuti dengan kluster Nusa Tenggara Timur (NTT) dan Gorontalo dengan jarak sama yang masing-masing membentuk kluster sendiri dengan anggota tunggal yaitu dirinya.

Dendogram metode *complete linkage* dengan jarak *euclid*, awalnya membentuk 5 kluster yaitu (Sumbar dan Jambi), (Sumsel dan Yogya), (Jatim dan Kaltim), (Banten dan Kalbar) dan (Sulsel dan Sultra). Kluster terakhir yang terbentuk yaitu Lampung.

Pada metode *average linkage* juga terbentuk lima kluster pada awal tahapannya, yaitu pada skala 0. Anggota kluster tersebut sama dengan metode *complete linkage* sedangkan kluster yang terakhir terbentuk sama dengan metode *single linkage*.

Dendogram metode ward juga menyatakan bahwa Lampung adalah kluster terakhir yang terbentuk dengan anggota tunggal yaitu dirinya. Pada metode median, kluster yang terakhir terbentuk sama dengan metode lainnya yaitu Lampung pada skala 25 diikuti Gorontalo pada skala 5-10 dan NTT pada skala 0-5. Pada awal tahapannya, metode ini membentuk empat kluster dengan salah satu kluster memuat anggota yang banyak. Pada akhir tahapan metode *centroid*, kluster yang terbentuk sama dengan metode *average linkage*. Kluster awal yang terbentuk ada 5 yaitu (Sumbar, Jambi, Bengkulu, NAD, Sumut, Banten, Kalbar, Jatim, Kaltim, Sumsel, Yogya, Jabar, Jateng, Jakarta dan Babel), (Riau dan Kalteng), (NTB dan Sulut), (Maluku dan Maluku Utara) dan (Sulsel, Sultra dan Kalsel).

Dari dendogram semua metode hirarki dapat dilihat bahwa metode-metode tersebut memberikan hasil yang mirip. Adapun provinsi yang selalu tergabung dalam satu kluster pada awal tahapannya atau provinsi-provinsi yang paling mirip yaitu (Sumbar dan Jambi), (Jatim dan Kaltim), (Sulsel dan Yogya), (Banten dan Kalbar) dan (Sumsel dan Sultra). Sedangkan provinsi yang paling tidak mirip dengan provinsi lain atau kluster yang terakhir terbentuk karena jaraknya paling jauh yaitu provinsi Lampung diikuti NTT dan Gorontalo.

Pada metode *single linkage*, kluster terkecil yang dapat dibentuk adalah 5 kluster. Anggota kluster tersebut adalah (Lampung), (Gorontalo), (NTT), (Maluku dan Maluku Utara) dan (provinsi-provinsi selain 4 kluster sebelumnya).

Metode *complete linkage* dapat membentuk 3, 4 atau 5 kluster. Jika diambil 4 maka anggotanya yaitu adalah (Lampung), (Sulsel, Sultra dan Kalsel), (NTB, Sulut, NTT, Maluku, Maluku Utara, Bali, Sulteng, dan Gorontalo) dan (provinsi-provinsi selain 3 kluster sebelumnya). Jika diambil 5 adalah pemecahannya maka anggota kluster tersebut adalah (Lampung), (Sulsel, Sultra dan Kalsel), (NTB, Sulut dan NTT), (Maluku, Maluku Utara, Bali, Sulteng, dan Gorontalo) dan (provinsi-provinsi selain 4 kluster sebelumnya).

Metode *average linkage* juga dapat membentuk 3, 4 atau 5 klaster. jika yang diambil 3 klaster adalah pemecahannya maka anggota klaster tersebut adalah (Lampung), (NTT) dan (provinsi-provinsi selain 2 klaster sebelumnya). Jika diambil 4 klaster maka anggotanya yaitu adalah (Lampung), (NTT), (Gorontalo) dan (provinsi-provinsi selain 3 klaster sebelumnya). Jika diambil 5 klaster adalah pemecahannya maka anggota klaster tersebut adalah (Lampung), (NTT), (Gorontalo), (Sulsel, Sultra dan Kalsel) dan (provinsi-provinsi selain 4 klaster sebelumnya).

Metode median dapat memberikan 4, 5 atau 6 hasil pengklasteran. Jika diambil 4 klaster maka anggota adalah (Lampung), (Gorontalo), (Sulsel, Sultra dan Kalsel) dan (provinsi-provinsi selain 3 klaster sebelumnya). Jika diambil 5 klaster maka anggotanya sama dengan metode *average linkage* yaitu (Lampung), (Gorontalo), (Sulsel, Sultra dan Kalsel), (NTT) dan (provinsi-provinsi selain 4 klaster sebelumnya). Jika diambil 6 klaster maka anggotanya (Lampung), (Gorontalo), (Sulsel, Sultra dan Kalsel), (NTT), (Nusa Tenggara Barat (NTB), Sulut, Maluku, Maluku Utara, Bali dan Sulteng) dan (provinsi-provinsi selain 5 klaster sebelumnya).

Dendogram metode ward mirip dengan metode *complete linkage* namun anggota klaster yang terbentuk tidak sama. Jarak yang terbentuk antar pemecahan tiap klaster lebih besar dibandingkan metode lainnya. Metode ini juga dapat membentuk 3, 4 atau 5 klaster. Jika diambil 3 klaster maka anggotanya yaitu (Lampung), (NTB, Sulut, NTT, Maluku, Maluku Utara, Bali, Sulteng dan Gorontalo) dan (provinsi-provinsi selain 2 klaster sebelumnya). Jika diambil 4 klaster maka anggotanya yaitu (Lampung), (NTB, Sulut, NTT, Maluku, Maluku Utara, Bali, Sulteng dan Gorontalo), (Sulsel, Sultra dan Kalsel) dan (provinsi-provinsi selain 3 klaster sebelumnya). Jika diambil 5 klaster maka anggotanya sama dengan metode *complete linkage*.

Pada metode *centroid*, jika diambil 3 atau 4 klaster adalah pemecahan yang ideal maka anggotanya sama dengan metode *average linkage*. Jika diambil 5 maka anggotanya sama dengan metode median.

Dari hasil dendogram dapat dilihat bahwa perubahan dari 3 klaster menjadi 4 klaster atau dari 4 klaster menjadi 5 klaster adalah penggabungan atau pemecahan provinsi yang sudah ada, bukan mengacak provinsi-provinsi dari awal.

Metode Pengklasteran *K-means*

Bila pengguna analisis klaster tidak mampu secara tegas menentukan berapa nilai k, banyaknya klaster maka disarankan untuk melakukan analisis komponen utama (Anonim, 2003c). Dengan menggunakan Minitab 13 *for window* diperoleh ada 5 komponen utama yang terbentuk.

Seperti telah disebutkan pada bab sebelumnya, bahwa metode ini dimulai dengan menentukan terlebih dahulu jumlah klaster yang diinginkan. Penulis menginginkan 5 klaster sesuai dengan analisis komponen utama yang telah dilakukan. Nilai *centroid* ditentukan oleh *software* secara acak. Dengan SPSS versi 11.0 *for window* diperoleh :

Tabel 4 *Cluster Membership*

| <i>Case Number</i> | <i>Provinsi</i> | <i>Cluster</i> | <i>Distance</i> |
|--------------------|-----------------|----------------|-----------------|
| 1 | NAD | 4 | 7,660 |
| 2 | Sumut | 4 | 8,143 |
| 3 | Sumbar | 4 | 4,905 |
| 4 | Riau | 4 | 11,696 |
| 5 | Jambi | 4 | 6,115 |
| 6 | Sumsel | 4 | 7,612 |
| 7 | Bengkulu | 4 | 8,319 |
| 8 | Lampung | 1 | ,000 |
| 9 | Babel | 4 | 10,365 |
| 10 | Jakarta | 4 | 6,921 |
| 11 | Jabar | 4 | 10,098 |
| 12 | Jateng | 4 | 8,412 |
| 13 | Yogya | 4 | 8,677 |
| 14 | Jatim | 4 | 6,283 |
| 15 | Banten | 4 | 5,477 |
| 16 | Bali | 3 | 11,677 |
| 17 | NTB | 3 | 7,580 |
| 18 | NTT | 3 | 17,732 |
| 19 | Kalbar | 4 | 5,990 |
| 20 | Kalteng | 4 | 12,111 |
| 21 | Kalsel | 5 | 7,014 |
| 22 | Kaltim | 4 | 3,651 |
| 23 | Sulut | 3 | 7,409 |
| 24 | Sulteng | 2 | 7,335 |
| 25 | Sulsel | 5 | 3,610 |
| 26 | Sultra | 5 | 5,220 |
| 27 | Gorontalo | 2 | 7,335 |
| 28 | Maluku | 3 | 11,525 |
| 29 | M/Utara | 3 | 11,554 |
| 30 | Papua | 4 | 13,204 |

Hasil output diatas ternyata tidak sama dengan hasil metode-metode hirarki. Jika 5 dipilih sebagai pemecahan hasil klaster yang ideal maka hasil output diatas, 3 klaster diantaranya sama dengan 3 diantara 5 pemecahan klaster metode *complete linkage* atau metode ward. Jika dibandingkan dengan 5 klaster pada metode *average linkage*, metode median atau metode *centroid* yang sama hanya 2 klaster. Sedangkan bila dibandingkan dengan metode *single linkage* hanya 1 klaster yang sama yaitu klaster provinsi Lampung.

Interpretasi Profil dan Akses Validitas Klaster

Dari hasil pengklasteran metode hirarki dan metode pengklasteran *K-means*, dengan alasan *reliability*, dapat diambil kesimpulan bahwa 5 adalah pemecahan hasil pengklasteran. Hal ini sesuai dengan analisis komponen utama dan hasil metode *complete linkage* dan metode ward.

Oleh karena itu, penulis mengambil kesimpulan bahwa 5 adalah penyelesaian pengklasteran dengan anggota sebagai berikut:

Tabel 5 Anggota kluster

| 1 | 2 | 3 | 4 | 5 |
|---------|--------------------|-----------------|-------|-------------------|
| Lampung | Sulawesi Selatan | Maluku | NTB | NAD |
| | Sulawesi Tenggara | Maluku Utara | Sulut | Sumut |
| | Kalimantan Selatan | Bali | NTT | Sumbar |
| | | Sulawesi Tengah | | Jambi |
| | | Gorontalo | | Sumsel |
| | | | | Bengkulu |
| | | | | Babel |
| | | | | Jakarta |
| | | | | Jabar |
| | | | | Jateng |
| | | | | Yogya |
| | | | | Jatim |
| | | | | Banten |
| | | | | Kalimantan Barat |
| | | | | Papua |
| | | | | Kalimantan Tengah |
| | | | | Riau |
| | | | | Kalimantan Timur |

Nama atau label dapat dicari dengan menguraikan *centroid* masing - masing kluster pada tiap variabel. Penamaan kluster dan pengambilan kesimpulan pengklasteran bersifat subyektif dan tergantung tujuan penelitian. Berikut nilai *centroid* masing-masing kluster dan *centroid* rataan populasi.

Tabel 6 *Centroid*

| No | Variabel | <i>Centroid</i> Populasi | Klaster 1 | Klaster 2 | Klaster 3 | Klaster 4 | Klaster5 |
|----|----------------|--------------------------|-----------|-----------|-----------|-----------|----------|
| 1 | Panas | 40.2960 | 31.38 | 35.1233 | 50.9140 | 52.3667 | 36.6922 |
| 2 | Sakit Kepala | 14.9630 | 13.45 | 15.5733 | 18.7980 | 19.8000 | 13.1739 |
| 3 | Batuk | 46.3287 | 44.59 | 35.6233 | 47.5340 | 53.0967 | 46.7467 |
| 4 | Pilek | 44.2607 | 45.84 | 31.3967 | 36.1600 | 51.9667 | 47.2828 |
| 5 | Diare | 4.9267 | 3.54 | 4.9833 | 5.3860 | 6.8933 | 4.5389 |
| 6 | Sesak Napas | 3.5937 | 2.11 | 2.6633 | 4.3940 | 3.9700 | 3.5461 |
| 7 | Asma | 3.2433 | 2.00 | 2.8467 | 3.4320 | 3.2200 | 3.3300 |
| 8 | Sakit Gigi | 6.8187 | 5.90 | 6.0467 | 8.8360 | 6.3800 | 6.5111 |
| 9 | Kejang | 0.9370 | 0.31 | 0.9967 | 1.7100 | 2.0433 | 0.5628 |
| 10 | Lumpuh | 0.7173 | 0.45 | 0.9433 | 0.9460 | 0.8867 | 0.6028 |
| 11 | Telinga Berair | 0.3210 | 0.24 | 0.2633 | 0.2900 | 0.5100 | 0.3122 |
| 12 | Campak | 2.3543 | 52.00 | 0.5133 | 0.6040 | 0.4867 | 0.7006 |
| 13 | Sakit Kuning | 0.5777 | 0.83 | 0.4900 | 0.7220 | 1.1100 | 0.4494 |
| 14 | Pikun | 1.1633 | 1.50 | 1.0100 | 1.2280 | 1.1800 | 1.1494 |
| 15 | Kecelakaan | 0.8470 | 0.67 | 1.4500 | 0.9760 | 0.4167 | 0.7922 |
| 16 | Lainnya | 25.3477 | 29.32 | 27.8300 | 22.8260 | 25.7667 | 25.3439 |

Pada klaster pertama (provinsi Lampung) terdapat nilai variabel keluhan kesehatan masyarakat yang mencolok yaitu campak dengan nilai 52,00 dari persentase penduduk yaitu 22,01 yang mempunyai keluhan kesehatan. Nilai ini sangat besar bila dibandingkan dengan *centroid* populasi dan klaster lain. Dibandingkan klaster lain, keluhan kesehatan lainnya, selain yang merupakan penyakit dan kecelakaan juga tinggi dibandingkan *centroid* populasi. Selain 2 variabel tersebut, sakit kuning, pikun dan pilek juga diatas *centroid* populasi. Nilai variabel lain, masih di bawah nilai *centroid* populasi.

Dari ciri-ciri diatas dapat dikatakan bahwa angka kesakitan (persentase penduduk yang mengalami keluhan kesehatan atau kesakitan) masih rendah. Dengan demikian klaster ini dapat diberi nama atau profil, klaster dengan angka kesakitan yang rendah.

Pada klaster 2, nilai *centroid* variabel yang berada diatas *centroid* populasi yaitu sakit kepala, diare, kejang, lumpuh, kecelakaan dan keluhan lainnya. Klaster ini dapat diberi nama klaster dengan angka kesakitan sedang.

Pada klaster 3, banyak terdapat nilai *centroid* variabel diatas *centroid* populasi yaitu ada 12 variabel. Variabel tersebut adalah panas, sakit kepala, batuk, diare, sesak napas, asma, sakit gigi, kejang, lumpuh, sakit kuning, pikun dan kecelakaan. Namun, variabel klaster ini tidak ada yang mencolok seperti nilai *centroid* campak pada klaster 1. Dari sini, dapat dilihat bahwa angka kesakitan pada klaster ini tinggi. Dengan demikian klaster ini dapat diberi nama, klaster dengan angka kesakitan tinggi.

Klaster 4 sama dengan klaster 3. Klaster ini juga mempunyai 12 nilai *centroid* variabel diatas rata-rata yaitu panas, sakit kepala, batuk, pilek, diare, sesak napas, kejang, lumpuh, telinga berair, sakit kuning, pikun dan keluhan lainnya. Namun, nilai-nilai pada variabel ini umumnya lebih tinggi dibandingkan klaster 3 bahkan jika dijumlahkan maka nilainya paling tinggi dibandingkan *centroid* populasi dan klaster lain. Dengan demikian klaster ini dapat diberi nama klaster dengan angka kesakitan yang lebih tinggi.

Angka kesakitan klaster ini lebih rendah dibandingkan klaster lain. Klaster ini hanya mempunyai 3 keluhan kesehatan yang nilainya diatas *centroid* populasi yaitu batuk, pilek dan asma. Selisih nilai batuk dan asma pun tidak jauh berbeda dengan nilai *centroid* populasi. Provinsi-provinsi pada klaster ini mempunyai derajat kesehatan yang secara umum lebih baik dibandingkan provinsi lain. Klaster ini mempunyai anggota yang lebih banyak bila dibandingkan klaster lain yaitu 18 provinsi. Dengan demikian, angka kesakitan bangsa Indonesia pada tahun 2003 masih dapat tergolong rendah. Kegagalan-kegagalan dibidang kesehatan dalam rangka meningkatkan derajat kesehatan masyarakat mungkin disebabkan faktor lain yaitu status gizi, umur harapan hidup dan angka kematian. Dari ciri-ciri diatas, maka klaster 5 dapat diberi nama klaster dengan angka kesakitan rendah sekali.

Akses validitas klaster secara tidak langsung telah dilakukan yaitu dengan melakukan anali sis klaster dengan metode yang berbeda dan membandingkan hasilnya.

KESIMPULAN DAN SARAN

Kesimpulan

1. Data persentase penduduk yang mempunyai keluhan kesehatan dalam sebulan pada tahun 2003 berdasarkan provinsi dan jenis keluhan dapat dibagi menjadi 5 klaster dengan rincian sebagai berikut:
 - i. Klaster 1 dengan profil klaster dengan angka kesakitan rendah. Klaster ini hanya mempunyai satu anggota yaitu provinsi Lampung
 - ii. Klaster 2 dengan profil klaster dengan angka kesakitan sedang. Anggota klaster ini yaitu provinsi Sulawesi Selatan, Sulawesi Tenggara dan Kalimantan Selatan
 - iii. Klaster 3 dengan profil klaster dengan angka kesakitan tinggi. Anggota klaster ini yaitu provinsi Maluku, Maluku Utara, Bali, Sulawesi Tengah dan Gorontalo.

- iv. Klaster 4 dengan profil klaster dengan angka kesakitan lebih tinggi. Anggota klaster ini yaitu provinsi NTB, NTT dan Sulawesi Utara.
 - v. Klaster 5 dengan profil klaster dengan angka kesakitan rendah sekali. Klaster ini mempunyai 18 anggota yaitu provinsi-provinsi selain provinsi yang ada pada 4 klaster sebelumnya. Provinsi tersebut adalah Nangro Aceh Darussalam, Sumatra Utara, Sumatra Barat, Riau, Jambi, Sumatra Selatan, Bengkulu, Bangka Belitung, Jakarta, Jawa Barat, Jawa Tengah, Yogya, Jawa Timur, Banten, Kalimantan Barat, Kalimantan Tengah, Kalimantan Timur dan Papua.
2. Secara umum, angka kesakitan masyarakat Indonesia pada tahun 2003 masih tergolong rendah. Kegagalan - kegagalan dibidang kesehatan, dalam rangka meningkatkan derajat kesehatan guna mensukseskan program pemerintah di bidang kesehatan Indonesia Sehat 2010 mungkin disebabkan faktor status gizi, umur harapan hidup dan angka kematian.
 3. Semua metode pengklasteran memberikan hasil yang mirip. Provinsi Lampung diikuti NTT dan Gorontalo selalu saja menjadi klaster terakhir yang terbentuk dan berbeda bila dibandingkan provinsi lain.
 4. Metode *complete linkage* lebih mirip dengan metode ward. Hasil metode *average linkage* sesuai dengan definisinya yang menggunakan seluruh informasi pada pasangan jarak. Metode metode median mirip dengan metode *centroid*.
 5. Hasil pengklasteran metode pengklasteran *K-means* untuk 5 klaster yang ideal lebih mirip dengan metode *complete linkage* atau metode ward yaitu 3 diantara klaster tersebut memberikan hasil anggota klaster yang sama. Bila dibandingkan dengan metode *average linkage*, metode median atau metode *centroid* yang hanya memberikan 2 klaster dengan anggota yang sama.
 6. Hasil pengklasteran yang lebih baik adalah dengan metode Ward atau metode *complete linkage* karena jarak yang terbentuk antar klaster lebih besar dibandingkan metode lainnya. Hal ini dapat dilihat pada dendogram (lampiran 6). Jarak yang besar menunjukkan klaster yang terbentuk memiliki nilai perbedaan yang besar pula.

Saran

1. Pemerintah dapat memberikan informasi (promotif) berupa penyuluhan kepada masyarakat dalam upaya mencegah penyakit yang datang terutama penyakit yang paling sering muncul dengan meningkatkan ketahanan diri dan menjaga kebersihan lingkungan (preventif).
2. Hasil masing-masing metode pengklasteran tidak sama persis, oleh karena itu akses validitas pengklasteran memang harus dilakukan.
3. Penamaan atau profil klaster sangat subyektif dan tergantung tujuan penelitian. Untuk dapat memberikan pengertian yang mendalam, diperlukan informasi atau fakta-fakta terbaru dalam bidang yang akan diteliti.

DARTAR PUSTAKA

- [1]. Anonim, 1991. *Informasi Ringkas Kesehatan*. Depkes RI. Jakarta.
- [2]. Anonim, 2002b. *Profil Kesehatan Provinsi Bengkulu 2001*, Dinas Kesehatan. Bengkulu.
- [3]. Anonim, 2003a. *Statistik Potensi Desa 2003*. CV Nasional. Jakarta.
- [4]. Anonim, 2004b. *Cluster Analysis*. <http://www.statsoft.com/textbook/stcluan.html>.
- [5]. Anonim, 2005b. *Cluster Analysis*. <http://149.170.199.144/multivar/ca.htm>.
- [6]. Anonim, 2005c. *Cluster Analysis*. <http://paleo.cortland.edu/class/stats/LectureNotes/11-cluster.pdf>.
- [7]. Anonim. 2002a. *Bengkulu Sehat 2010*. Dinkes Provinsi Bengkulu, Bengkulu.
- [8]. Anonim. 2003b. *Analisis Peubah Ganda*. Jurusan Statistika IPB, Bogor.
- [9]. Anonim. 2003c. *Modul Praktikum Pelatihan Analisis Multivariat*. Jurusan Statistika, Bogor.

- [10]. Anonim. 2004a. *Indonesia Sehat 2010*. <http://www.depkes.go.id>.
- [11]. Argadiredja, D. 2003. *Program Penanggulangan Kemiskinan Bidang Kesehatan*. <http://www.bppt.go.id/rakorbangnas03/depkes4.pdf>.
- [12]. Dillon, W.R. & M. Golstein. 1984. *Multivariate Analysis Method & Applications*. John Wiley & Sons, inc, Canada.
- [13]. Milne, G.R. 2005. *ClusterAnalysis*.
<http://intra.som.umass.edu/georgemilne/powerpoint/cluster%anal%ysis.ppt>
- [14]. Moloe, 1999. ----. <http://www.sehat2010.com>.
- [15]. Notoatmodjo, S. 2002. *Metodologi Penelitian Kesehatan*. Rineka Cipta. Jakarta.
- [16]. Portier, K.M. 2003. *Some Old and New Approaches to Cluster Analysis*. Florida. <http://www.ifasstat.ufl.edu/sta4702/pdf/lecture10.pdf>.
- [17]. Sambamoorthi, N. 2005a. *Hierarchical Cluster Analysis: Some Basic and Algorithms*. <http://www.crmportals.com>.
- [18]. Santoso, S. 2004. *Buku Latihan SPSS Statistik Multivariat*. Elex Media Komputindo, Jakarta.
- [19]. Seber, G.A.F. 1984. *Multivariate Observations*. John Wiley & Sons. New York. USA.
- [20]. Sudjana. 2002. *Metoda Statistika*. Tarsito. Bandung
- [21]. Supranto, J. 2004. *Analisis Multivariat: Arti dan Interpretasi*. Rineka Cipta. Jakarta.
- [22]. Venables, W.N. & B.D. Ripley. 1995. *Modern Applied Statistics with S-Plus*, Springer, Verlag. New York.
- [23]. Wichern, D.W. 2002. *Multivariate Statistical Analysis*. Prentice Hall, New Jersey.