ISSN: 2339-2541
JURNAL GAUSSIAN, Volume 8, Nomor 4, Tahun 2019, Halaman 530 - 541
Online di: http://ejournal3.undip.ac.id/index.php/gaussian

PERBANDINGAN NILAI KORELASI PADA KANONIK *ROBUST (METODE MINIMUM COVARIANCE DETERMINANT)* DAN KANONIK KLASIK (Studi Kasus Data Struktur Ekonomi dan Kesejahteraan Rakyat di Jawa Barat 2016)

Widi Rahayu[1], Sudarno[2], Alan Prahutama[3]
[1,2,3] Departemen Statistika FSM Universitas Diponegoro
dsghani@gmail.com

ABSTRACT

Canonical correlation analysis is a multivariate statistical analysis that aims to examine the correlation between two groups of variabels in a way to maximize the value of correlation between variabels. The outlier in the data affect the covariance matrix is generated, So that use robust multivarat. There is robust multivariate approach to the analysis of canonical robust with MCD method (Minimum Covariance Determinant). This final project aims to determine comparison between correlation value of robust canonical with MCD and canonical classical methods. With a data theres containing of outliers in the case studies of people's welfare and economic structures in West Java in 2016. Used a set of variabels welfare of people consist of 6 variabel (Y) and a set of variabels economic structure which consists of four variabels (X). Based on the analysis results obtained that robust canonical correlation values better explain the correlation between two sets of variabels, the correlation value r_{c1}0.99552, r_{c2}=0.91228, r_{c3}=0.71529, r_{c4}=0.63174, While the correlation value on classical canonical are r_{c1} =0.931489, r_{c2} =0.538672, r_{c3} =0.387099, r_{c4} =0.259318, Canonical robust can be interpreted more because it meets the test of significance are partially and directly, while the classical canon can not be interpreted further because it does not meet the test of the significance of the function.

Keywords : Classical canonical correlation, canonical correlation robust correlation value, Minimum Covariance Determinant (MCD)

1. PENDAHULUAN

Pada beberapa masalah penelitian statistika multivariat, ketertarikan peneliti tidak hanya pada pembentukan model regresi linear antara variabel dependen dan variabel independen saja, tetapi peneliti juga tertarik pada hubungan linear antara dua himpunan variabel. Analisis korelasi kanonik adalah jawaban dan penggunaan pada jenis masalah penelitian statistika multivariat seperti ini (Hair et al., 2010). Analisis ini berdasarkan pada matriks varian kovarian yang membentuk suatu kombinasi linier dari setiap kelompok variabel menjadi maksimum, namun matriks varian kovarian pada analisis ini sangat sensitif terhadap *outlier*, yang akan mempengaruhi matriks kovarian (Romanazzi, 1992).

Outlier dalam data multivariat dapat dideteksi melalui jarak mahalonobis dan melalui jarak *robust*. Rousseeuw dan Von Zomeren (1999) menjelaskan bahwa penggunaan jarak *robust* mampu meminimumkan pengaruh dari adanya efek *masking* dan *swamping* dalam pendeteksian *outlier*, jarak *robust* menggunakan penaksir dari rata-rata (μ) dan varian kovarian ($\sum$). *Minimum Covariance Determinant* (MCD) yang dikemukakan oleh Rousseeuw dan dibahas lebih lanjut oleh Hubert pada tahun 2008. MCD memilki matriks varian kovarian yang dihasilkan dan menjadi alternatif sebagai pengganti matriks varian kovarian pada korelasi kanonik klasik, mendeteksi seluruh *outlier* dalam data dan memiliki ketahanan yang cukup besar terhadap *outlier* (Riana, 2012).

2. TINJAUAN PUSTAKA

2.1. Analisis Korela si Kanonik

2.1.1 Pengertian dan Tujuan Analisis Korelasi Kanonik

Analisis korelasi kanonik adalah teknik analisis statistik yang digunakan untuk mengukur tingkat keeratan hubungan antara dua kelompok variabel. Analisis korelasi kanonik berfokus pada korelasi antara kombinasi linear dari himpunan variabel dependen $\mathbf{y}'$ = $(\mathbf{y_1}, \mathbf{y_2}, \ldots, \mathbf{y_p})$ dengan kombinasi linear dari himpunan variabel independen $\mathbf{x}'$ = $(\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_p})$. Ide dari analisis ini adalah menentukan pasangan dari kombinasi linear yang memiliki korelasi terbesar. Kemudian, mencari pasangan dari kombinasi linear di antara pasangan yang tidak berkorelasi pada pasangan bagian di awal yang dipilih. Pasangan dari kombinasi linear ini disebut fungsi kanonik, dan korelasinya disebut korelasi kanonik (Johnson dan Wichern, 2007).

Tujuan dari analisis korelasi kanonik pada dasarnya sama dengan korelasi sederhana, dimana menurut (Misbahuddin dan Hasan 2004) korelasi sederhana yaitu mengukur hubungan keeratan antara dua variabel, dan korelasi berganda yakni mengukur keertan antara tiga variabel atau lebih. Namun berbeda dengan korelasi sederhana jumlah variabel pada korelasi kanonik X dan Y lebih dari satu, sehingga analisis korelasi kanonik ini digolongkan pada multivariat.

2.1.2 Asumsi-Asumsi dalam Korelasi Kanonik

Menurut (Hair *et a*l., 2010) asumsi-asumsi pada analisis korelasi kanonik adalah:

a. Linieritas
b. Variabel independen dan variabel dependen berdistribusi normal multivariat.
c. Homoskedastisitas
d. Nonmultikolinieritas

2.2. Deteksi *Outlier* Multivariat

Outlier adalah kasus atau data yang memiliki karakteristik unik yang terlihat sangat jauh dari observasi-observasi lainnya dan muncul dalam bentuk nilai ekstrim baik pada data univariate maupun pada data multivariat (Ghozali, 2006). Terdapat beberapa cara untuk mendeteksi *outlier* dalam data multivariat, yaitu melalui jarak mahalanobis atau melalui jarak *Robust.*

2.2.1. Jarak Mahalonobis

Identifikasi *outlier* pada multivariat umumnya berdasarkan pada jarak Mahalanobis Johnson dan Wichern (1998) menyatakan bahwa pengamatan ke-i didefinisikan sebagai *outlier* jika jaraknya lebih besar dari nilai *chi-square* pada sejumlah variabel. Perhitungan tersebut sebagai berikut:

$$\boldsymbol{MD^2} = (\boldsymbol{x_i} - \boldsymbol{\mu})^{\boldsymbol{t}} \sum(\boldsymbol{x_i} - \boldsymbol{\mu}) \quad , i = 1,2, \ldots, n$$

dimana:

$\boldsymbol{\mu} =$ vektor rata-rata, $n =$ banyaknya pengamatan, $\sum$ = matriks varian kovarian

Sebuah pengamatan x_i diidentifikasikan sebagai *outlier* jika jarak mahalanobisnya lebih besar dari nilai *chi-square* tabel dengan tingkat signifikansi $(1-\alpha)$ dan derajat bebas $MD^2 > \chi_p^2$ atau setara dengan $MD > \sqrt{\chi_p^2}$.

Penggunaan jarak mahalanobis untuk pendeteksian *outlier* pada variabel multivariat menjadi tidak maksimal jika terdapat lebih dari satu pengamatan *outlier*, karena adanya pengaruh *masking* dan *swamping*.

2.2.2. Jarak *Robust*

Jarak *robust* merupakan suatu pendekatan untuk mengidentifikasi *outlier* pada data multivariat, dengan menggunakan penaksir dari rata-rata (μ) dan (varian kovarian) $\sum$ pada metode *robust*. Sehingga metode ini mampu meminimumkan pengaruh dari adanya efek *swamping* dan *masking* dalam pendeteksian *outlier* (Rencher, 2002).

Pendeteksian *outlier* dalam kasus data mutivariat dilakukan berdasarkan jarak kuadrat *robust* dari x_i ke T, yang didefinisikan sebagai berikut:

$$RD_i^2 = (x_i - \bar{\boldsymbol{t}}_{1j})' \sum_{mcd(1)}^{-1} (x_i - \bar{\boldsymbol{t}}_{1j}), i = 1,2,\dots,n$$

dimana:

$\bar{\boldsymbol{t}}_{1j}$ = Vektor rata-rata $\qquad$ n = Banyaknya pengamatan

$\sum_{mcd(1)}^{-1}$ = Matriks varian kovarian dari data $\boldsymbol{X}$ yang mempunyai determinan matriks terkecil

Sebuah pengamatan diidentifikasikan sebagai *outlier* jika jarak *robust* nya lebih besar dari nilai *chi-square* dengan tingkat signifikansi (1-α) dan derajat bebas = p

$RD^2 > \chi_p^2$ atau setara dengan $RD > \sqrt{\chi_{p)}^2}$.

2.3. Penduga Koefisien Korelasi Kanonik

Misal hubungan antara himpunan variabel dependen $y_1, y_2, \dots, y_p$ yang dinotasikan dengan vektor acak **y**, dengan himpunan variabel independen X1, $x_2, \dots, x_q$ yang dinotasikan dengan vektor acak **x,** dimana p≤q. Maka tiap vektor observasi di dalam suatu sampel dipartisi menjadi :

$$\begin{bmatrix} y \\ \dots \\ x \end{bmatrix} = \begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{ip} \\ \cdots \\ x_{i1} \\ x_{i2} \\ \vdots \\ x_{iq} \end{bmatrix}, \text{I} = 1, 2, \dots, \text{n}$$

Untuk tiap sampel pada *n* vektor observasi, maka vektor rata-rata dan matriks kovariannya :

$$\begin{bmatrix} \bar{y} \\ \dots \\ \bar{x} \end{bmatrix} = \begin{bmatrix} \bar{y}_1 \\ \bar{y}_2 \\ \vdots \\ \bar{y}_p \\ \cdots \\ \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_q \end{bmatrix}$$

Dimana :
ȳ = mean sampel variabel dependen
$\bar{x}$ = mean sampel variabel independen
$\bar{y}_p$= matriks mean partisi ke-p variabel dependen
$\bar{x}_q$ = matriks mean partisi ke-q variabel independen
P = banyak variabel dependen
q= banyak variabel independen

Matriks varian kovarian
sampel **S** = (s_{jk}) adalah matriks sampel varian kovarian dari *p* variabel:

$$\mathbf{S} = \begin{bmatrix} s_{11} & s_{12} & \dots s_{1p} \\ s_{21} & s_{22} & \dots s_{2p} \\ \vdots & \vdots & \ddots \ \vdots \\ s_{p1} & s_{p2} & \dots s_{pp} \end{bmatrix}$$

Pada matriks **S,** varian sampel dari *p* variabel terletak pada diagonal utama, dan pasangan kovarian sampel berada diluar diagonal utama.

Varian sampel pada variabel ke-*j* adalah $s_{jj} = s_j^2 s_{jj} = s_j^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_{ij-}\bar{y}_j)^2$

Varian sampel pada variabel ke-j dan ke- k , $s_{jk}\ s_{jk} = \frac{1}{n-1}\sum_{i=1}^{n}(y_{ij-}\bar{y}_j)(y_{ik-}\bar{y}_k)$

Varian s_{jj} dinyatakan sebagai s_j^2, kuadrat dari standar deviasi s_j dan **S** simetris $s_{jk} = s_{kj}$.

Setiap sampel pada *n* vektor observasi, maka matriks varian dan kovariannya adalah:

$$\mathbf{S} = \begin{bmatrix} \mathbf{S_{yy}} & \mathbf{S_{yx}} \\ \mathbf{S_{xy}} & \mathbf{S_{xx}} \end{bmatrix}$$

Dapat dilihat pada persamaan berikut untuk menjelaskan persamaan sebelumnya misal p = 3 dan q = 3

Maka :

$$\begin{bmatrix} \bar{y} \\ \bar{x} \end{bmatrix} = \begin{bmatrix} \bar{y}_1 \\ \bar{y}_2 \\ \bar{y}_3 \\ \bar{x}_1 \\ \bar{x}_2 \\ \bar{x}_3 \end{bmatrix};$$

Untuk membuat matriks varian dan kovarian sampel dengan p=3 dan q=3 adalah :

dimana:
Matriks varian kovarian sampel dari p variabel
$\mathbf{S_{yy}}$ = Varian sampel pada variabel dependen y
$\mathbf{S_{yx}}$ = Varian sampel pada variabel dependen y dan variabel independen x
$\mathbf{S_{xy}}$ = Varian sampel pada variabel independen y dan variabel dependen x
$\mathbf{S_{xx}}$ = Varian sampel pada variabel independen x

ukuran matriksnya adalah:
$\mathbf{S_{yy}}$ = p x p
$\mathbf{S_{yx}}$ = p x q
$\mathbf{S_{xy}}$ = q x p
$\mathbf{S_{xx}}$ = q x q
S = Berukuran Simetris

$$\mathbf{S} = \begin{bmatrix} \mathbf{S_{yy}} & \mathbf{S_{yx}} \\ \mathbf{S_{xy}} & \mathbf{S_{xx}} \end{bmatrix} \qquad \mathbf{S} = \begin{bmatrix} s_{y1}^2 & s_{y1y2} & s_{y1y3} & s_{y1x1} & s_{y1x2} & s_{y1x3} \\ s_{y2x1} & s_{y2x2} & s_{y2x3} & s_{y2x1} & s_{y2x2} & s_{y2x3} \\ s_{y3x1} & s_{y3x2} & s_{y3}^2 & s_{y3x1} & s_{y3x2} & s_{y3x3} \\ s_{x1y1} & s_{x1y2} & s_{x1y3} & s_{x1}^2 & s_{x1x2} & s_{x1x3} \\ s_{x2y1} & s_{x2y2} & s_{x2y3} & s_{x2x1} & s_{x2}^2 & s_{x2x3} \\ s_{x3y1} & s_{x3y2} & s_{x3y3} & s_{x3x1} & s_{x3x2} & s_{x3}^2 \end{bmatrix}$$

(Rencher, 2002)

Kombinasi linear dari kedua himpunan variabel dapat dituliskan :

$U = \boldsymbol{a^T y} = \boldsymbol{a_1^T y} + \boldsymbol{a_2^T y} + \cdots + \boldsymbol{a_k^T y}$ Untuk : $\mathrm{Var}(U) = \boldsymbol{a^T}\mathrm{Cov}\,(\mathbf{y})\,\mathbf{a} = \boldsymbol{a^T S_{yy} a}$

$V = \boldsymbol{b^T x} = \boldsymbol{b_1^T x} + \boldsymbol{b_2^T x} + \cdots + \boldsymbol{b_k^T y}$ $\qquad \mathrm{Var}(U) = \boldsymbol{b^T}\mathrm{Cov}\,(\mathbf{y})\,\mathbf{b} = \boldsymbol{b^T S_{xx} a}$

Dimana :
U = Variabel kanonik untuk variabel dependen
V = Variabel kanonik untuk variabel independen
a=Vektor koefisien variabel dependen
b=Vektor koefisien variabel

Korelasi digunakan untuk mengetahui keeratan hubungan linier antara masing-masing variabel, cara yang dapat digunakan adalah dengan menghitung matriks korelasi antar semua variabel (Sembiring, 2003).

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

Dengan $r_{jk} = \dfrac{\sum_{i=1}^{n}(y_{ik}-\bar{y}_j)(y_{ik}-\bar{y}_k)}{\sqrt{\sum_{i=1}^{n}(y_{ik}-\bar{y}_j)^2}\sqrt{\sum_{j=1}^{n}(y_{ik}-\bar{y}_k)^2}}$

Untuk $j = 1, 2, \ldots, p$ dan $k = 1, 2, \ldots, p$ dan $r_{jk} = r_{kj}$ untuk semua j dan k.
(Johnson and Wichern, 2007)

Nilai r_{jk} berada antara, $-1 \leq r_{ik} \leq +1$ ketika $r_{ik} = 0$ maka artinya tidak ada hubungan linier antar komponen, hubungan linier sempurna bila $r_{ik} = \pm 1, \neq 1$ artinya hubungannya searah dan -1 bila berlawanan arah (Sembiring, 2003).

Sehingga korelasi kanonik :

$r_{c(u,v)} = \frac{\text{cov }(U,V)}{\sqrt{\text{Var}(U)\text{Var}(V)}} = \frac{\mathbf{a}^T S_{yx}\mathbf{b}}{\sqrt{\mathbf{a}^T S_{yy}\mathbf{b}}\sqrt{\mathbf{b}^T S_{yy}\mathbf{b}}}$

Vektor koefisien **a** dapat diperoleh dengan cara mencari $\lambda_1 > \lambda_2 > \ldots > \lambda_k$ yang merupakan nilai eigen dari matriks $\mathbf{S}_{yy}^{-1/2}\mathbf{S}_{yx}\mathbf{S}_{xx}^{-1}\mathbf{S}_{xy}\mathbf{S}_{yy}^{-1/2}$ yang berpadanan dengan vektor eigen $\mathbf{a_1}, \mathbf{a_2}, \ldots, \mathbf{a_k}$. Misal $\mathbf{b_1}, \mathbf{b_2}, \ldots, \mathbf{b_k}$ merupakan vektor eigen dari matriks $\mathbf{S}_{xx}^{-1/2}\mathbf{S}_{xy}\mathbf{S}_{yy}^{-1}\mathbf{S}_{yx}\mathbf{S}_{xx}^{-1/2}$ dimana nilai λ_1 pertama diperoleh dari $\boldsymbol{b_k} = \mathbf{S}_{xx}^{-1/2}\mathbf{S}_{xy}\mathbf{S}_{yy}^{-1/2}\mathbf{a}_k$, dengan k = 1, 2, …, λ (Johnson and Wichern, 2007).

Nilai eigen dapat diperoleh dari persamaan karakteristik :

$|\mathbf{S}_{yy}^{-1}\mathbf{S}_{yx}\mathbf{S}_{xx}^{-1}\boldsymbol{S_{xy}} - \lambda\mathbf{I}| = \mathbf{0}$, dan $|\mathbf{S}_{xx}^{-1}\mathbf{S}_{xy}\mathbf{S}_{yy}^{-1}\mathbf{S}_{yx} - \lambda\mathbf{I}| = \mathbf{0}$

Koefisien vektor $\mathbf{a_k}$ dan $\mathbf{b_k}$ diperoleh pada fungsi kanonik $\mathbf{U_k} = \mathbf{a_k^T y}$ dan $\boldsymbol{V_k} = \mathbf{b_k^T x}$ adalah vektor eigen dari dua matriks yang sama :

$(\mathbf{S}_{yy}^{-1}\mathbf{S}_{yx}\mathbf{S}_{xx}^{-1}\boldsymbol{S_{xy}} - \lambda\mathbf{I})\mathbf{a} = 0$, dan $(\mathbf{S}_{xx}^{-1}\mathbf{S}_{xy}\mathbf{S}_{yy}^{-1}\boldsymbol{S_{yx}} - \lambda\mathbf{I})\mathbf{b} = \mathbf{0}$

Sehingga dua matriks $\mathbf{S}_{yy}^{-1}\mathbf{S}_{yx}\mathbf{S}_{xx}^{-1}\boldsymbol{S_{xy}}$ dan $\mathbf{S}_{xx}^{-1}\mathbf{S}_{xy}\mathbf{S}_{yy}^{-1}\boldsymbol{S_{yx}}$ mempunyai nilai eigen tak nol dan berbeda vektor eigen (Rencher, 2002).

Untuk pasangan fungsi kanonik ke-*k* :

$$\begin{array}{ll} U_1 = \mathbf{a_1^T y} & \mathbf{V_1} = \mathbf{b_1^T x} \\ \vdots & \vdots \\ U_k = \mathbf{a_k^T y} & \mathbf{V_k} = \mathbf{b_k^T x} \end{array}$$

Dimana **y** dan **x** adalah nilai-nilai dari himpunan variabel dependen dan independen untuk satuan pengamatan khusus. Secara umum, terdapat *k* = min (*p,q*) nilai dari korelasi kanonik r_{ck} dengan *k* pasang variabel kanonik yang bersesuaian $U_k = \mathbf{a_k^T y}$ dan $V_k = \mathbf{b_k^T x}$. Sebagai contoh, jika *p* = 3 dan *q* = 7, maka terdapat tiga korelasi kanonik r_{c1}, r_{c2}, dan r_{c3}.

$$\text{Corr}_{(U_k, V_k)} = \frac{\mathbf{a_k^T S_{yx} b}}{\sqrt{\mathbf{a_k^T S_{yy} a_k}}\sqrt{\mathbf{b_k^T S_{xx} b_{kyy}}}} = r_{ck}$$ (Johnson and Wichern, 2007)

2.4.Uji Hipotesis

Ada dua hipotesis yang akan diujikan dalam analisis korelasi kanonik yaitu uji korelasi kanonik secara keseluruhan dan uji secara sebagian (Rencher, 2002).

1. Uji Korelasi Kanonik secara Keseluruhan

Hipotesis :

H_0 : $r_{c2} = r_{c2} = \ldots = r_{ck} = 0$ (semua korelasi kanonik tidak signifikan)

H_1 : $r_{ci} > 0$ (paling tidak ada satu korelasi kanonik signifikan, dengan i = 1, 2,…k)

Tingkat signifikansi : $\alpha = 5\%$

Statistik Uji :

$$F = \frac{1-\Lambda_1^{1/t}}{\Lambda_1^{1/t}}\frac{df_2}{df_1}$$

dengan : $\Lambda_1 = \prod_{i=1}^{k}(1 - r_i^2)$

$df_1 = pq$

$df_2 = wt - \frac{1}{2}pq + 1$

$w = n - \frac{1}{2}(p + q + 3)$

$t = \sqrt{\frac{p^2q^2 - 4}{p^2 + q^2 - 5}}$

2. Uji Secara Sebagian

Hipotesis :

H_0 : $r_{cj} = 0$ (korelasi kanonik tidak signifikan)

H_1 : $r_{cj} > 0$ (paling tidak ada satu korelasi kanonik signifikan)

Tingkat signifikansi : $\alpha = 5\%$

Statistik Uji :

$$F = \frac{1-\Lambda_j^{1/t}}{\Lambda_j^{1/t}} \frac{df_2}{df_1}$$

dengan : $\Lambda_j = \prod_{i=j}^{k}(1 - r_i^2)$

$df_1 = (p - j + 1)(q - j + 1)$

$df_2 = wt - \frac{1}{2}[\{p - j + 1)(q - j + 1)] + 1$

$w = n - \frac{1}{2}(p + q + 3)$

$$t = \sqrt{\frac{(p-j+1)^2\,(q-j+1)^2 - 4}{(p-j+1)^2 + (q-j+1)^2 - 5}}$$

dimana :
n = jumlah pengamatan
p = banyak himpunan variabel **y**
q = banyak himpunan variabel **x**
Kriteria penolakan :
H_0 ditolak jika F hitung > $F_{\alpha;df1;df2}$ atau $\Lambda_1 \leq \Lambda_{\alpha,p,q,n-1-q}$ dimana :
n = jumlah pengamatan
p = banyak himpunan variabel **y**
q = banyak himpunan variabel **x**
Kriteria penolakan :
H_0 ditolak jika F hitung > $F_{\alpha;df1;df2}$ atau $\Lambda_1 \leq \Lambda_{\alpha,p,q,n-1-q}$
(Rencher, 2002)

2.5. Analisis Redundansi

Redundansi merupakan sebuah nilai (ukuran) yang menunjukkan besarnya keragaman yang dapat dijelaskan berdasarkan korelasi antara variabel dependen dan independen dengan variabel kanonik.
Besarnya keragaman untuk himpunan **y** yang diterangkan oleh $V_1, V_2, \ldots, V_k$:

$$R(\boldsymbol{y}|V) = \frac{\sum_{i=1}^{p} R_{Yi}^2\,|\mathbf{x}}{p}$$

Indeks Redundansi **y** yang diterangkan oleh $V_1, V_2, \ldots, V_k$:

$$RI(\boldsymbol{y}|V) = R(\boldsymbol{y}|V)r_{ck}^2$$

Besarnya keragaman untuk **x** yang diterangkan oleh $U_1, U_2, \ldots, U_k$:

$$R(\boldsymbol{x}|U) = \frac{\sum_{i=1}^{p} R_{Xi}^2\,|\mathbf{y}}{q}$$

Indeks Redundansi **x** yang diterangkan oleh $U_1, U_2, \ldots, U_k$:

$$RI(\boldsymbol{x}|U) = R(\boldsymbol{x}|U)r_{ck}^2$$

Untuk mendapatkan indeks redundansi yang tinggi, salah satunya harus memiliki korelasi kanonik yang tinggi dan derajat yang tinggi dari keragaman bersama yang dapat dijelaskan oleh tiap variabel kanonik. Indeks redundansi dihitung untuk kedua variabel dependen dan independen (Hair *et al.*, 2010).

2.6.Interpretasi Fungsi Kanonik

Interpretasi yang dapat dilakukan dalam analisis korelasi kanonik yaitu terhadap koefisien kanonik (bobot kanonik / *weight* kanonik), *loadings* kanonik dan *cross loadings* kanonik. Bobot kanonik merupakan koefisien kanonik yang telah dibakukan, dapat diinterpretasikan sebagai besarnya kontribusi variabel asal terhadap variate kanonik. Semakin besar nilai koefisien ini menyatakan semakin besar kontribusi variabel yang bersangkutan terhadap variate kanonik.

3. METODOLOGI PENELITIAN

3.1. Jenis dan Sumber Data

Data yang akan digunakan dalam penelitian ini adalah data sekunder struktur ekonomi dan kesejahteraan rakyat Jawa Barat per Kabupaten/Kota Tahun 2016 yang diperoleh dari Perpustakaan Badan Pusat Statistik Jawa Barat, Buku Kabupaten dalam Angka 2017.

3.2. Variabel Penelitian

Variabel X yang digunakan dalam penelitian ini adalah struktur ekonomi sesuai dengan yang tertera pada buku kabupaten dalam angka 2017 sebagai berkut:

X_1: Persentase PDRB dari sektor pertanian.
X_2: Persentase pekerja dari sektor pertanian.
X_3: Persentase dari jenis pekerjaan utama 1 (tenaga professional. teknisi dan yang sejenis), atau 2 (tenaga kepemimpinan dan ketatalaksanaan), atau (tenaga usaha dan yang sejenis).
X_4: Persentase dengan status pekerja utama sebagai pekerja keluarga.

Variabel Y yang digunakan dalam penelitian ini adalah kesejahteraan rakyat sesuai dengan yang tertera pada buku kabuaten dalam angka 2017 sebagai berkut:

Y_1: Persentase penduduk dengan pengeluaran di atas UMR perkapita per bulan
Y_2: Persentase rumah tangga dengan penerangan listrik/petromak
Y_3: Persentase rumah yang memiliki TV/Video/VCD
Y_4: Persentase rumah tangga dengan jenis bahan bakar (minyak tanah/kayu/gas) untuk memasak
Y_5: Persentase penduduk dengan pendidikan terakhir SMA atau Perguruan Tinggi
Y_6: Persentase Angka Kelahiran Total

3.3. Software yang digunakan

Software yang digunakan dalam analisis data pada tugas akhir ini adalah *SPSS 23,* dan *R 3.4.3.*

3.4. Metode Analisis

Langkah analisis data pada tugas akhir ini adalah sebagai berikut:

1. Menguji asumsi korelasi kanonik sebagai berikut: linieritas, variabel dependen dan independen berdistribusi normal multivariat, homoskedastisitas, dan nonmultikolinieritas.
2. Pendeteksian *outlier* pada keseluruhan data pengamatan menggunakan jarak mahalanobis dan juga menggunakan jarak *robust*.
3. Menyusun matriks varian kovarian dengan metode MCD menggunakan software excel dan software R.3.4.3 untuk mendapatkan matriks pada metode korelasi kanonik *robust* dengan menggunakan metode MCD .
4. Menyusun matriks varian kovarian untuk metode korelasi kanonik menyusun Matriks varian kovarian **S** dipakai apabila data yang diolah memiliki satuan yang sama.
5. Mencari vektor eigen berdasarkan nilai eigen yang telah diperoleh dengan persamaan $\left(\mathbf{S}_{\mathrm{yy}}^{-1}\mathbf{S}_{\mathrm{yx}}\mathbf{S}_{\mathrm{xx}}^{-1}\boldsymbol{S}_{\boldsymbol{xy}} - \lambda\mathbf{I}\right)\mathbf{a} = \mathbf{0}$ dan $(\mathbf{S}_{\mathrm{xx}}^{-1}\mathbf{S}_{\mathrm{xy}}\mathbf{S}_{\mathrm{yy}}^{-1}\mathbf{S}_{\mathrm{yx}} - \lambda\mathbf{I}|\mathbf{b} = \mathbf{0}$
6. Vektor eigen **a** dan **b** yang merupakan koefisien variabel kanonik dan bobot kanonik.
7. Menentukan nilai penduga koefisien korelasi kanonik dan fungsi kanonik banyaknya fungsi kanonik terbentuk mengikuti minimal banyak variabel dalam setiap variat.
8. Menguji signifikansi korelasi kanonik secara keseluruhan dan secara sebagian. Jika fungsi kanonik tidak signifikan, maka hubungan antar variabel tidak akan diinterpretasikan.
9. Analisis redundansi variabel kanonik, ukuran untuk besar persentase keragaman yang dijelaskan oleh variabel kanonik.
10. Membandingkan hasil terbaik diantara korelasi kanonik dengan metode MCD dan korelasi kanonik klasik.

4. HASIL DAN PEMBAHASAN

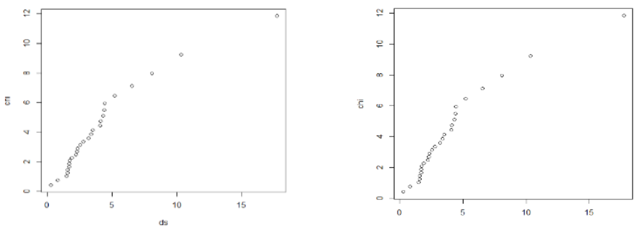
4.1. Uji Asumsi Korelasi Kanonik

a. Uji Linieritas

Hasil perhitungan dengan SPSS 23 diperoleh hasil Sig (.000) < α(0.05) yang berarti terdapat hubungan linier antara variable X dan Variabel Y.

b. Distribusi Normal Multivariat

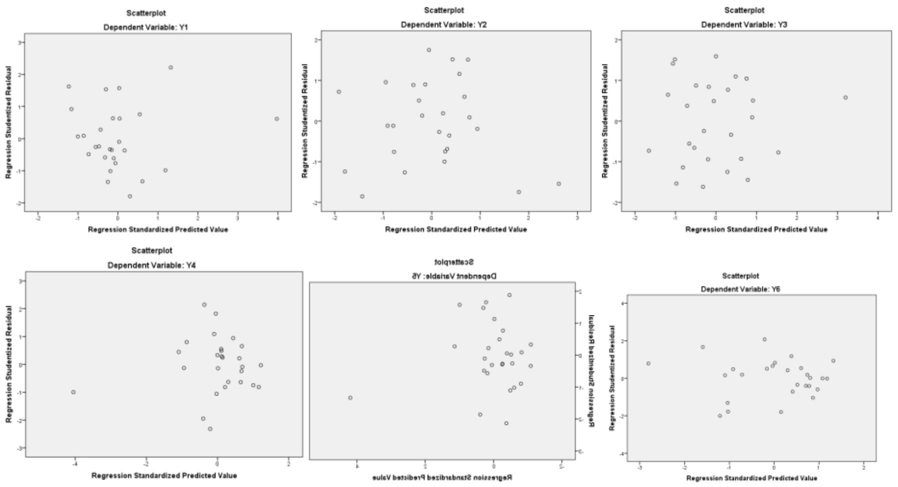
Hasil perhitungan R 3.4.3 diperoleh *p-value* > α(0.05) yang berarti variable independen dan dependen berdstribusi normal multivariat. Adapun terdapat setengah atau lebih nilai ${d_j}^2 \leq q_{c,0,5,4=}3{,}35669$. Secara grafis data variable independen dan variable dependen dapat dikatakan berdistribusi normal multivariat, sebab plot diantara ${d_j}^2$ yang telah diurutkan dengan kuantil ${\chi_j}^2$ mendekati bentuk garis lurus.



Gambar 1. Grafik Chi-Square Asumsi Normal Multivariat pada Variabel Dependen dan Independen

c. Homoskedastisitas

Pengujian secara visual data memperlihatkan adanya suatu sebaran tidak sistematis antara residual kuadrat dengan penjelas x sehingga asumsi homoskedastisitas terpenuhi.



Gambar 2. Scatterplot antara y_1, y_2, y_3, y_4, y_5, y_6 terhdap x_1, x_2,x, x_4

d. Nonmultikolinieritas

Hasil perhitungan dengan SPSS 23 diperoleh nilai *Tolerance value* > 0.10 *dan VIF* < 10 yang berarti variable dependen dan independen tidak terjadi mulltikolinieritas.

4.2.Penentuan Koefsien Korelasi Kanonik Robust

Dengan menggunakan software ms. Excel 2010, R 3.4.3 diperoleh:

Fungsi	$\lambda_i = r_{ci}^2$	Koefisien korelasi kanonik	Presentase	Presentase Kumulatif
1	0.991068	0.995524	36.24882	36.24882
2	0.832257	0.912281	30.44022	66.68904
3	0.511646	0.715294	18.7137	85.40274
4	0.399099	1. 0.631743	14.59726	100

Fungsi ke-1 mengakomodasi 36.2488% hubungan kanonik, sedangkan 30.4402 diakomodasi dalam fungsi ke-2, 18.7137% diakomodasi dalam fungsi ke-3, sedangkan pada fungsi ke-4 yaitu sebesar 14.59726% . Korelasi kanonik pada fungsi ke-1 sebesar 0.991068 lebih besar dibanding korelasi kanonik pada fungsi ke-2.

4.3 Penentuan Koefsien Korelasi Kanonik Klasik

Dengan menggunakan software ms. Excel 2010, R 3.4.3 diperoleh:

Fungsi	$\lambda_i = r_{ci}^2$	Koefisien korelasi kanonik	Presentase	Presentase Kumulatif
1	0.991068	0.995524	63.24882	63.24882
2	0.832257	0.912281	30.44022	66.68904
3	0.511646	0.715294	18.7137	85.40274
4	0.399099	0.631743	14.59726	100

4.3.Fungsi Korelasi Kanonik Robust

$U_1 = -0.0631Y_1 + 0.0284Y_2 + 0.06848Y_3 - 0.54322Y_4 - 0.83391Y_5 + 0.00227Y_6$

$V_1 = 0.11276X_1 + 0.00509X_2 + 0.95297X_3 + 0.28125X_4$

Pasangan fungsi kanonik ke-2

$U_2 = -0.10152Y_1 - 0.08383Y_2 + 0.07983Y_3 + 0.98618Y_4 - 0.0610Y_5 - 0.00167Y_6$

$V_2 = [-0.06077X_1 + 0.07402X_2 - 0.98382X_3 + 0.15135X_4]$

Pasangan Fungsi Kanonik ke-3

$U_3 = 0.00042\,Y_1 - 0.04311Y_2 - 0.01773Y_3 - 0.9292Y_4 - 0.3666Y_5 + 0.00125\ Y_6$

$V_3 = [0.18650X_1 - 0.50432X_2 - 0.54092X_3 + 0.64673X_4]$

Pasangan Fungsi ke-4

$U_4 = 0.34761Y_1 + 0.35560Y_2 - 0.10879Y_3 + 0.84821Y_4 - 0.14604Y_5 + 0.00824Y_6$

$V_4 = [-0.69140\,X_1 - 0.03343X_2 + 0.24245X_3 + 0.67974X_4]$

4.4.Fungsi Korelasi Kanonik Robust

Pasangan Fungsi ke-1

$U_1 = 0.27641Y_1 + 0.17273Y_2 - 0.57597Y_3 - 0.66211Y_4 + 0.67233Y_5 + 0.03027Y_6$

$V_1 = 0.34403X_1 - 0.99184X_2 - 0.5701X_3 + 0.10867X_4$

Pasangan fungsi kanonik ke-2

$U_2 = 0.61399Y_1 - 0.21367Y_2 - 0.12652Y_3 - 0.09643Y_4 + 0.96667Y_5 + 0.02482Y_6$

$V_2 = [0.05501X_1 + 0.10107X_2 + 0.12599X_3 + 0.98685X_4]$

Pasangan Fungsi Kanonik ke-3

$U_3 = -0.87047\,Y_1 + 0.46366Y_2 - 0.39006Y_3 - 0.77616Y_4 - 0.15079Y_5 + 0.10005\,Y_6$

$V_3 = [0.26547X_1 + 0.7087X_2 - 0.9900X_3 + 0.11899X_4]$

Pasangan Fungsi Kanonik ke-4

$U_4 = 0.34761Y_1 + 0.35560Y_2 - 0.10879Y_3 + 0.84821Y_4 - 0.14604Y_5 + 0.00824Y_6$

$V_4 = [0.99904X_1 + 0.31715X_2 + 0.27577X_3 - 0.12338X_4]$

j	**Df_1**	**Df_2**	**Kanonik *Robust***				**Kanonik Klasik**			
			Fhitung	F-tabel	Keputusan	Kesimpulan	Fhitung	F-tabel	Keputusan	Kesimpulan
1	24	60.3	3.27	1.7001	H_0 ditolak	r_{c1} signifikan	9.46053	1.700	H_0 ditolak	r_{c2} tidak signifikan
2	15	48	0.532	1.8801	H_0 diterima	r_{c2} tidak signifikan	-2.8680	1.746	H_0 diterima	r_{c2} tidak signifikan
3	8	38	-3,8980	2,1935	H_0 diterima	r_{c3} tidak signifikan	-4,4347	2.193	H_0 diterima	r_{c3} tidak signifikan
4	3	8.6	1.75634	4.0661	H_0 diterima	r_{c4} tidak signifikan	-2.7068	4.476	H_0 diterima	r_{c4} tidak signifikan

4.5.Uji Signifikansi Sebagian Korelasi Kanonik *Robust*

4.6.Uji Signifikansi Keseluruhan Korelasi Kanonik *Robust*

a. Hipotesis

H_0 : r_{c1}=r_{c2} =r_{c3} =r_{c4} $= 0$ (semua korelasi kanonik *robust* tidak signifikan)

H_1 : $r_{ci} \neq 0$ (paling tidak ada korelasi kanonik signifikan, dengan i= 1,2,3,4)

b. Tingkat Signifikansi

α = 5%

c. Statistik Uji

$F = \frac{1-\Lambda_j^{1/t} df_2}{\Lambda_j^{1/t} df_1} = 3{,}27$

$\Lambda = \prod_{i=1}^{4}(1-r_i^2)\Lambda =$ 0,00044

d_{f1}= (4-1+1)(6-1+1) =24

d_{f2}=wt-$\frac{1}{2}pq + 1$= 60.34

w= n – $\frac{1}{2}(p+q+3)$=20.5

t=$\sqrt{\frac{p^2q^2-4}{p^2+q^2-5}}$= 3.48

d. Daerah Penolakan

Ho ditolak karena F=3.27 > $F_{0.05(24,60,34)}$=1.70017

e. Keputusan

Nilai F hitung adalah 3.27 yang nilainya lebih dari $F_{0.05(24,60,34)}$ = 1.70017, sehingga Ho ditolak pada tingkat signifikansi 5%. Kesimpulan

f. Kesimpulan

Pada tingkat signifikansi 5% nilai F hitung adalah 3.27 menunjukan bahwa fungsi korelasi kanonik *robust* tidak bernilai nol atau ada korelasi yang signifikan.

4.7.Uji Signifikansi Keseluruhan Korelasi Kanonik *Klasik*

a. Hipotesis

H_0 : r_{c1}=r_{c2}=r_{c3}=r_{c4}=0(Semua korelasi kanonik klasik tidak signifikan)

H_1 : $r_{ci} \neq 0$ (paling tidak ada satu korelasi kanonik klasik signifikan dengan i = 1,2,3,4)

b. Tingkat Signifikansi

α = 5%

c. Statistik Uji

$F = \frac{1-\Lambda_j^{1/t} df_2}{\Lambda_j^{1/t} df_1} = -0.5674$

$\Lambda = \prod_{i=1}^{4}(1-r_i^2) = 0.074485$

d_{f1}= (4-1+1)(6-1+1) =24

d_{f2}=wt-$\frac{1}{2}pq + 1$= 60.34

w= n – $\frac{1}{2}(p+q+3)$=20.5

t=$\sqrt{\frac{p^2q^2-4}{p^2+q^2-5}}$= 3.48

d. Daerah Penolakan

Ho diterima karena F=-0.56747 <$F_{0.05(24,58,51596)}$=1.700117

e. Keputusan Nilai F hitung adalah -0.56747 yang nilainya lebih dari $F_{0.05(24,58,51596)}$ = 1.700117, sehingga Ho diterima pada tingkat signifikansi 5%.

f. Kesimpulan Pada tingkat signifikansi 5% Nilai F hitung adalah -0,56747 38 menunjukan bahwa fungsi korelasi kanonik klasik tidak signifikan.

4.8.Analisis Redundansi Korelasi Kanonik *Robust*

Fungsi 1	Muatan kanonik	Muatan Kanonik (Kuadrat)	Rata-Rata Muatan Kanonik	Korelasi Kanonik r_{c1}^2	Indeks Redundansi
Y1	0.60022	0.360264			
Y2	0.10005	0.01001			
Y3	0.21439	0.045963			
Y4	0.58285	0.339714			

Y5	0.76934	0.591884			
Y6	0.12235	0.01497			
Variabel dependen		**1.362805**	**0.227134**	**0.991068**	**0.22509**
X1	0.18325	0.033581			
X2	0.30498	0.093013			
X3	0.82906	0.68734			
X4	0.17541	0.030769			
Variabel Independen		**0.844703**	**0.21117**	**0.991068**	**0.209289**

Indeks redudansi pada fungsi kanonik pertama untuk variabel dependen (Kesejahteraan rakyat) menjelaskan 22.5% besarnya keragaman variabel kanonik pada variabel independen (struktur ekonomi). Demikian juga, pada fungsi kanonik pertama variabel independen (struktur ekonomi) menjelaskan 20.9% besarnya keragaman variabel kanonik pada variabel dependen (kesejahteraan rakyat). Tidak dilakukan interpretasi pada s korelasi kanonik *robust* yang ke-2,3 dan 4 tidak signifikan sehingga tidak diinterpretasikan

4.9.Analisis Redundansi Korelasi Kanonik Klasik

Karena pada uji signifikansi korelasi kanonik klasik secara keseluruhan dan sebagian didapat bahwa korelasi kanonik klasik tidak ada yang signifikan secara statistik maka fungsi tidak dapat diinterpretasikan lebih lanjut.

4.10. Interpretasi Korelasi Kanonik

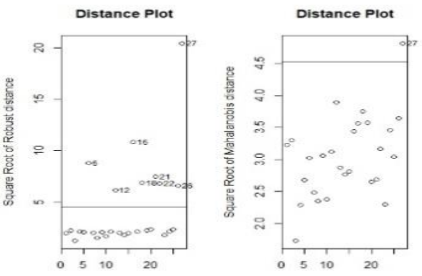
Interpretasi fungsi kanonik yang digunakan adalah bobot kanonik, berdasarkan koefisien fungsi kanonik yang telah dibakukan pada setiap variabel kanonik (pada table *standardized canonical cofficiente for dependen* variabel *function* dan *standardized canonical cofficiente for Covariate*) didapat :

$$\begin{bmatrix} y1 \\ y2 \\ y3 \\ y4 \\ y5 \\ y6 \end{bmatrix} = \begin{bmatrix} -0.18065 & 0.18447 & 0.9595 & 1.35551 \\ 0.33787 & 0.98067 & -0.15563 & 0.49061 \\ 0.42711 & 0.08206 & 1.23515 & -0.99676 \\ 0.26014 & -0.36864 & 0.82816 & 0.31147 \\ -0.18588 & -0.56473 & 0.22431 & -0.63371 \\ -0.38438 & 0.05579 & -1.17254 & -0.16422 \end{bmatrix} \qquad \begin{bmatrix} x1 \\ x2 \\ x3 \\ x4 \end{bmatrix} = \begin{bmatrix} -0.3514 & 0.0858 & -1.07242 & 0.89744 \\ 0.01441 & -0.73285 & -0.64823 & -1.93208 \\ -1.39317 & -0.38606 & -0.21299 & -0.25702 \\ -0.53988 & -0.56988 & 1.24050 & 1.20688 \end{bmatrix}$$

Untuk variabel dependen, pada fungsi kanonik *robust* pertama urutan kontribusi relative terhadap variabell kanonik dependen adalah y_3 yaitu sebesar 0.42711, y_2 yaitu sebesar 0.33787. y_4 yaitu sebesar 0.26014, y_1 yaitu sebesar -0.18065, y_6 yaitu sebesar -0.38438, y_5 yaitu sebesar -1.18588. Untuk variabel independen pada fungsi kanonik *robust* pertama urutan kontributisi relative terhadao variabel kanonik dependen adalah x_2 yaitu sebesar 0.01441, x_1 yaitu sebesar -0.3514, x_4 yaitu sebesar -0.53988, dan x_3 yaitu sebesar -1.39317.

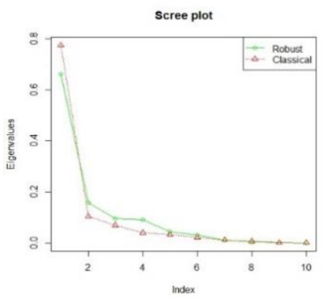
4.11. Perbandingan Korelasi Kanonik *Robust* dengan Menggunakan *Metode Minimum Covariance Determinant* dan Kanonik Klasik

Plot Jarak Mahalanobis dan Jarak *Robust*



Gambar 3. Perbandingan Plot Jarak Mahalanobis dan Jarak *Robust*

4.7.1 Grafik Perbandingan Nilai Eigen



Gambar 4. Perbandingan nilai eigen analisis korelasi kanonik *robust* menggunalan metode *Minimum Covariance Determinance* dan analisis korelasi kanonik klasik.

5. PENUTUP

5.1. Kesimpulan

1. Pada pengujian asumsi-asumsi korelasi kanonik, semua asumsi korelasi kanonik: linieritas, variabel dependen dan independen berdistribusi normal multivariat, homoskedastisitas, dan nonmultikolinieritas terpenuhi.
2. Berdasarkan pengolahan dan pembahasan data secara manual maupun dengan bantuan *software* didapat koefisien korelasi kanonik *robust* dengan nilai korelasi r_{c1}= 0.99552, $r_{c2} = 0.91228$, $r_{c3} = 0.71529$, $r_{c4} = 0.63174$. Sedangkan nilai korelasi pada kanonik klasik $r_{c1} = 0.931489$, $r_{c2} = 0.538672$, $r_{c3} = 0.387099$, $r_{c4} = 0.259318$, dimana nilai korelasi *robust* pada kasus ini lebih besar jika dibandingkan dengan korelasi klasik.
3. Pada uji signifikansi korelasi kanonik secara keseluruhan dan sebagian dengan hasil hanya korelasi kanonik *robust* pertama signifikan maka fungsi korelasi kanonik *robust* pertama dapat diinterpretasi secara lanjut yaitu dilakukan analisis redundansi. Pada uji signifikansi korelasi kanonik secara keseluruhan dan sebagian didapat bahwa korelasi kanonik klasik pertama, kedua, ketiga, dan keempat tidak signifikan secara statistik sehingga tidak dapat dilakukan nterpretasi lebih lanjut.

DAFTAR PUSTAKA

Ghozali, I. 2006. *Aplikasi Analisis Multivariat dengan Program SPSS*. Semarang: Universitas Diponegoro.

Hair, *et al*. 2010. *Multivariat Data Analysis, Seventh Edition.* New Jersey: Pearson Prentice Hall International Inc.

Johnson, R. dan Wichern, D. (2007). *Applied Multivariat* Statistik*a Analysis.* United States of America: Pearsonn Prentice Hall.

Misbahuddin. dan Hasan, I. (2004). *Analisis Data Penelitian dengan Statistik.* Edisi kedua. Jakarta: PT Bumi Aksara.

Rencher, A.C. 2002. *Methods of Multivariat Analysis*. Canada: John Wiley & Sons.

Riana, F. 2012. *Perbandingan Metode Kekar Biweight Midcovariance dan Minimum Covariance Determinant dalam Analisis Korelasi Kanonik.* Bogor: Tesis Program Studi Statistika Terapan. Institut Pertanian Bogor.

Romanazzi M. 1992. Influence in Canonical Correlation Analysis.*Psychometrika*.57:237-259http://www.springerlink.com/content/l24p13843114jr65/.

Rousseeuw PJ, Van Driessen K. 1999. A Fast Algorithm for the Minimum Covariance Determinant Estimator. *Technometrics Journal*. 3:212-223.

Sembiring, R. 2003. *Analisis Regresi.* Bandung: Penerbit ITB.