

---

## Analisis Diskriminan Linear Robust Dengan Metode Winsorized Modified One-Step M-Estimator

Mega Selvia Tjahaya<sup>1</sup>, Raupong<sup>2</sup>, Georgina Maria Tinungki<sup>3\*</sup>  
<sup>123</sup>Departemen Statistika, Fakultas MIPA,  
Universitas Hasanuddin, Makassar, 90245, Indonesia  
email: meaatj21@gmail.com

### Abstract

Discriminant analysis is a method used to classify an individual (object) into a group. Discriminant analysis is divided into classical linear discriminant analysis and classical quadratic discriminant analysis. Discriminant analysis must fulfilled the assumptions of normality and homogeneity of the variance-covariance matrix, however this method is very sensitive to data contains outliers. Robust linear discriminant analysis with the winsorized modified one-step M-estimator(WMOM) approach is a method that can resolve outliers data. WMOM works by trimming these outliers then replacing the outliers with the highest or lowest value of the remaining data by using criteria trimming MOM. This study aims to obtain a linear robust discriminant function with the WMOM method using the Sn scale on diabetes and prediabetes data for the period December 2016-January 2017. Based on the results of the analysis and discussion of this method, the discriminant function is obtained with a classification error rate of 16.67%.

**Keywords:** Diabetes, One-Step M-estimator, Prediabetes, Robust Linear Discriminant Analysis, Winsorized Modified.

### Abstrak

Analisis diskriminan merupakan suatu metode yang digunakan untuk mengklasifikasikan suatu individu (objek) ke dalam suatu kelompok. Analisis diskriminan terbagi atas analisis diskriminan linear klasik dan analisis diskriminan kuadrat klasik. Analisis diskriminan harus memenuhi asumsi kenormalan dan kehomogenan matriks varians-kovarians, namun metode ini sangat sensitif terhadap data yang mengandung pencilan. Analisis diskriminan linear robust dengan pendekatan winsorized modified one-step M-estimator (WMOM) merupakan metode yang dapat mengatasi data pencilan. WMOM bekerja dengan cara memangkas pencilan tersebut kemudian mengganti nilai pencilan dengan nilai tertinggi atau terendah dari data yang tersisa dengan menggunakan kriteria pemangkasan Modified One-Step M-Estimator (MOM). Penelitian ini bertujuan untuk mendapatkan fungsi diskriminan linear robust dengan metode WMOM menggunakan skala Sn pada data diabetes dan prediabetes Periode Desember 2016 - Januari 2017. Berdasarkan hasil analisis dan pembahasan dari metode ini diperoleh fungsi diskriminan dengan tingkat kesalahan klasifikasi sebesar 16,67%.

**Kata kunci:** Analisis Diskriminan Linear Robust, Diabetes, Pradiabetes, Winsorized Modified One-Step M-estimator.

## 1. Pendahuluan

Analisis diskriminan merupakan suatu metode statistika yang digunakan untuk mengklasifikasikan suatu individu atau objek ke dalam suatu kelompok yang telah ditentukan sebelumnya berdasarkan variabel-variabel bebasnya[1]. Analisis yang biasa digunakan adalah analisis diskriminan klasik meliputi analisis diskriminan linear klasik dan analisis kuadrat klasik. Terdapat dua asumsi yang harus dipenuhi dalam analisis

diskriminan linear klasik yaitu variabel bebas berdistribusi normal multivariat dan matriks kovariansi dari dua kelompok data yang diamati adalah sama.

Analisis diskriminan klasik tidak dapat bekerja dengan baik jika data yang dianalisis mengandung pencilan dikarenakan rata-rata sampel dan matriks varian kovarian sampel sangat sensitive terhadap pencilan sehingga diperlukan suatu penaksir yang *robust*. Analisis diskriminan yang menggunakan penaksir yang *robust* disebut sebagai analisis diskriminan *robust*. Ada berbagai penduga robust dan telah diadopsi untuk menangani pencilan dalam data yang tidak dapat ditangani oleh pendekatan analisis diskriminan konvensional.

Salah satu penduga yang digunakan untuk mengatasi masalah pencilan adalah *Modified One-Step M-Estimator* (MOM). Modifikasi *One-Step M-Estimator* adalah penduga robust dengan breakdown point yang tinggi yang melakukan *trimming* (memangkas proporsi nilai ekstrim) sehubungan dengan distribusi secara empiris menentukan apakah suatu observasi harus dipangkas atau tidak, serta berapa jumlah pemangkasan yang diperlukan[2].

Karena fokus dari pekerjaan ini adalah untuk memperoleh analisis yang lebih akurat, penduga MOM juga memanfaatkan *Winsorized Modified One-Step M-Estimator* (WMOM). WMOM melakukan penggantian nilai-nilai terbesar dan terkecil yang kontinu dari data setelah *trimmed* menggunakan kriteria MOM sehingga kemungkinan kehilangan informasi akibat pemangkasan dapat dihindari. Modifikasi tertentu akan dibuat ketika mengadopsi WMOM, yang mencakup penduga skala *robust* ( $S_n$ ) dan matriks kovarian winsorized untuk mendapatkan fungsi analisis diskriminan pada penderita diabetes dan prediabetes di Puskesmas Kecamatan Pesanggrahan Kota Jakarta periode Desember 2016 - Januari 2017.

## **2. Material dan Metode**

Data yang digunakan pada penelitian ini adalah data sekunder. Data sekunder yang digunakan adalah data penderita diabetes dan prediabetes diperoleh dari Puskesmas Kecamatan Pesanggrahan Kota Jakarta Selatan Provinsi DKI Jakarta periode Desember 2016 - Januari 2017 [3]. Data ini terdiri dari 90 sampel pengamatan dan tujuh variabel prediktor dengan dua kelompok.

Variabel yang digunakan dalam penelitian ini adalah:

1. Kategori pengklasifikasian terdiri dari dua kelompok yaitu kelompok 1 sebagai kelompok prediabetes, kelompok 2 sebagai kelompok diabetes.
2. Variabel ( $X$ ) yang terdiri dari faktor-faktor yang mempengaruhi kejadian diabetes. Variabel yang akan digunakan, yaitu Usia ( $X_1$ ), Berat Badan ( $X_2$ ), Kolesterol ( $X_3$ ), Trigliserida ( $X_4$ ), Tekanan Darah Sistolik ( $X_5$ ), Tekanan Darah Diastolik ( $X_6$ ), Asam Urat ( $X_7$ ).

Untuk menentukan metode yang digunakan adalah analisis diskriminan klasik atau robust haruslah dilakukan pengujian asumsi terhadap kenormalan, homogenitas matriks varian kovarian dan keberadaan pencilan pada analisis data

### 2.1. Asumsi

#### 1. Uji Kesamaan Vektor Rata-rata

Pengujian terhadap vektor rata-rata dilakukan untuk melihat apakah terdapat perbedaan yang nyata antar kelompok yang terbentuk dari setiap variabel prediktor. Pengujian ini dilakukan dengan hipotesis:

$$H_0: \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \dots = \boldsymbol{\mu}_k$$

$$H_1 : \text{ada } \boldsymbol{\mu}_g \neq \boldsymbol{\mu}_q \text{ untuk } g \neq q \text{ dengan } g = q = 1, 2, \dots, k$$

Statistik uji yang digunakan adalah statistik V-Bartlett yang mengikuti distribusi *chi-square* dengan derajat bebas  $p(k - 1)$ . Statistik V-Bartlett dilakukan dengan:

$$V = - \left[ (N - 1) - \frac{p + k}{2} \right] \ln(\Lambda) \quad (1)$$

dengan

$$\Lambda = \frac{|\mathbf{W}|}{|\mathbf{B} + \mathbf{W}|} = \frac{\left| \sum_{g=1}^k \sum_{i=1}^{n_g} (\mathbf{X}_{gi} - \bar{\mathbf{X}}_g)(\mathbf{X}_{gi} - \bar{\mathbf{X}}_g)^t \right|}{\left| \sum_{g=1}^k \sum_{i=1}^{n_g} (\mathbf{X}_{gi} - \bar{\mathbf{X}})(\mathbf{X}_{gi} - \bar{\mathbf{X}})^t \right|} \quad (2)$$

dengan :

$\Lambda$  = Wilk's Lamda

$\mathbf{W}$  = matriks jumlah kuadrat dan hasil kali data dalam kelompok.

$\mathbf{B}$  = matriks jumlah kuadrat dan hasil kali data antar kelompok.

$n_g$  = jumlah pengamatan kelompok ke-g.

$N$  = jumlah pegamatan

$p$  = jumlah variabel prediktor

$k$  = jumlah kelompok.

Apabila  $V \leq \chi_{\alpha, p(k-1)}^2$ , maka  $H_0$  diterima, sebaliknya jika  $V > \chi_{\alpha, p(k-1)}^2$ , maka  $H_0$  ditolak [4].

#### 2. Uji Normalitas Multivariat

Pengujian asumsi normal multivariat dapat dilakukan dengan menghitung ukuran jarak Mahalanobis pada setiap pengamatan dan nilai *Chi - Square*[5]. Hipotesis yang akan digunakan adalah sebagai berikut.

$$H_0 : \text{Data berdistribusi normal multivariat}$$

$$H_1 : \text{Data tidak berdistribusi normal multivariat}$$

Uji normalitas ini dilakukan dengan membuat plot jarak Mahalanobis ( $d_i^2$ ) dan distribusi Chi-Square  $\left( \chi_{\frac{1}{n}(i-0,5), p}^2 \right)$ . Langkah-langkahnya sebagai berikut.

- 1) Menghitung jarak Mahalanobis dengan persamaan (7)

- 2) Setiap  $d_i^2$  akan mengikuti sebaran *Chi-Square*  $\chi_{\frac{1}{n}(i-0,5),p}^2$ . Dengan  $\chi_{\frac{1}{n}(i-0,5),p}^2$  adalah distribusi *Chi-Square* dengan  $p$  adalah banyaknya variabel prediktor.
- 3) Mengurutkan nilai  $d_i^2$  dari nilai  $d_i^2$  terkecil sampai  $d_i^2$  terbesar. Plot *Chi-Square* akan memeriksa apakah  $d_i^2$  mengikuti sebaran *Chi-Square* atau tidak.
- 4) Membuat plot dengan titik koordinat  $(d_i^2; \chi_{\frac{1}{n}(i-0,5),p}^2)$ .
- 5)  $H_0$  diterima jika lebih dari 50% nilai  $d_i^2 \leq \chi_{\frac{1}{n}(i-0,5),p}^2$ , ini berarti data berdistribusi normal multivariat. Jika lebih kecil dari 50% maka  $H_0$  ditolak.

### 3. Uji Homogenitas Matriks Variansi Kovariansi

Pada analisis diskriminan linear matriks variansi kovariansi antar kelompok pada seluruh variabel prediktor homogen. Jika tidak maka akan dilakukan analisis diskriminan kuadratik. Pengujian kesamaan matriks variansi kovariansi antar kelompok akan menggunakan statistik uji *Box's M*. Untuk menguji kesamaan matriks variansi kovariansi ( $\Sigma$ ) antar kelompok, digunakan hipotesis :

$$H_0 : \Sigma_1 = \Sigma_2 = \dots = \Sigma_k$$

$$H_1 : \text{ada } \Sigma_g \neq \Sigma_q \text{ untuk } g \neq q \text{ dengan } g = 1, 2, \dots, k$$

Statistik Uji :

$$C = (1 - u)M \tag{3}$$

dengan

$$M = \left[ \sum_{g=1}^k (n_g - 1) \right] \ln |\mathbf{S}_{gab}| - \sum_{g=1}^k [(n_g - 1) \ln |\mathbf{S}_g|] \tag{4}$$

$$\mathbf{S}_{gab} = \frac{\sum_{g=1}^k (n_g - 1) \mathbf{S}_g}{\sum_{g=1}^k (n_g - 1)} \tag{5}$$

$$u = \left[ \sum_{g=1}^k \frac{1}{(n_g - 1)} - \frac{1}{\sum_{g=1}^k (n_g - 1)} \right] \left[ \frac{2p^2 + 3p - 1}{6(p + 1)(k - 1)} \right] \tag{6}$$

dengan

$n_g$  = jumlah pengamatan kelompok ke-g.

$p$  = jumlah variabel prediktor.

$\mathbf{S}_g$  = matriks variansi kovariansi kelompok ke-g.

$\mathbf{S}_{gab}$  = matriks variansi kovariansi gabungan

$k$  = jumlah kelompok

Jika  $C > \chi_{\alpha, \frac{1}{2}(p(p+1)(k-1))}^2$  yang berarti terdapat perbedaan matriks variansi kovariansi. Uji *Box's M* bekerja dengan baik jika  $n_g$  lebih dari 20 dan jika  $p$  dan  $k$  kurang dari lima [5].

## 2.2. Pencilan

Identifikasi data pencilan pada data multivariat umumnya menggunakan jarak Mahalanobis. Jarak mahalanobis adalah ukuran jarak yang didasarkan pada korelasi antar variabel-variabel, khususnya invers matriks kovariansi. Pengamatan ke- $i$  didefinisikan sebagai data pencilan multivariat jika jarak mahalanobisnya ( $d_i^2$ ) lebih besar dari nilai  $chi-square(\chi_{\alpha,p}^2)$  pada  $p$  buah variabel. Jarak mahalanobis ditentukan dengan rumus

$$d_i^2 = (\mathbf{X}_i - \boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1} (\mathbf{X}_i - \boldsymbol{\mu}) \quad (7)$$

dengan  $\boldsymbol{\mu}$  dan  $\boldsymbol{\Sigma}$  menyatakan rata-rata vektor dan matriks variansi kovariansi.

WMOM- $S_n$  adalah salah satu estimator *robust* untuk mengestimasi vektor rata-rata dan matriks variansi kovariansi yang digunakan untuk mengestimasi jarak Mahalanobis sehingga jarak Mahalanobis *robust* ( $d_{i,WMOM-S_n}^2$ ) dari ke- $i$  diidentifikasi sebagai pencilan jika

$$d_{i,WMOM-S_n}^2 = (\mathbf{X}_i - \boldsymbol{\mu}_{WMOM-S_n})^t \boldsymbol{\Sigma}_{WMOM-S_n}^{-1} (\mathbf{X}_i - \boldsymbol{\mu}_{WMOM-S_n}) > \chi_{\alpha,p}^2 \quad (8)$$

$\boldsymbol{\mu}_{WMOM-S_n}$  dan  $\boldsymbol{\Sigma}_{WMOM-S_n}$  menyatakan vektor rata-rata dan matriks variansi kovariansi yang diestimasi dengan metode WMOM- $S_n$ .

## 2.3. Analisis Diskriminan Linear

Analisis diskriminan digunakan dalam menguji variabel apa yang membedakan antara kelompok secara jelas. Analisis diskriminan berfokus pada bobot dan hubungan dengan setiap variabel prediktor. Variabel-variabel dengan bobot diskriminan tertinggi akan menunjukkan perbedaan terkuat pada kelompok [6].

Model dasar analisis diskriminan adalah sebuah persamaan yang menunjukkan suatu kombinasi linear dari berbagai variabel prediktor, yaitu:

$$D = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + \dots + b_pX_p \quad (9)$$

dengan  $D$  = Skor diskriminan linear;  $\mathbf{b}$  = koefisien diskriminan ;  $\mathbf{X}$ = variabel prediktor.

## 2.4. Modified One-Step M-Estimator

*Modified One-Step M-Estimator* (MOM) secara empiris menentukan apakah pengamatan harus dipangkas atau tidak (*trimmed*) [7]. Secara matematik, Wilcox (2002) mendefinisikan estimator MOM sebagai [7]

$$\hat{\theta}_j = \frac{\sum_{i=i_1+1}^{n_j-i_2} X_{ij}}{n_j - i_1 - i_2} \quad (10)$$

dengan

$X_{ij}$  = pengamatan urutan ke- $i$  pada variabel prediktor ke- $j$

$\hat{M}_j$  = Med( $X_j$ ) median dari data pada setiap variabel ke- $j$

$i_1$  = banyaknya  $X_{ij}$  yang memenuhi  $(X_{ij} - \hat{M}_j) < -2.24$  (*skala estimator*)

$i_2$  = banyaknya  $X_{ij}$  yang memenuhi  $(X_{ij} - \hat{M}_j) > 2.24$  (*skala estimator*)

$n_j$  = jumlah data pada setiap variable [7]

### 2.5. Winsorized Modified One-Step M-Estimators

Pada WMOM, observasi yang dipangkas diganti dengan nilai tertinggi dan terendah dari data yang tersisa. Pada dasarnya, WMOM mengikuti pendekatan *trimmed* yang mempertimbangkan bentuk distribusi data selama pemangkasan. Pemangkasan ini memberikan perhatian lebih ke pusat daripada ujung distribusi data. Hanya pencilan yang akan dipangkas melalui pendekatan *trimmed* otomatis ini. Namun, nilai *trimmed* akan diganti dengan data tertinggi dan terendah yang tersisa, bukan hanya menghilangkannya. Masalah kehilangan informasi karena proses *trimmed* dapat dikurangi karena winsorisasi selalu mempertahankan ukuran sampel asli [8].

Konstruksi sampel Winsorized dilanjutkan sebagai berikut. Untuk setiap variabel acak  $X_j = \{x_{1j}, \dots, x_{mj}\}$ ,  $j = 1, \dots, p$  sampel winsorized diperoleh dari

$$w_{ij} = \begin{cases} x_{(i_1+1)j} & \text{jika } x_{ij} \leq x_{(i_1+1)j} \\ x_{ij} & \text{jika } x_{(i_1+1)j} < x_{ij} < x_{(m+i_1)j} \\ x_{(m+i_2)j} & \text{jika } x_{ij} \geq x_{(m+i_2)j} \end{cases} \quad (11)$$

dimana

$i_1$  = Jumlah data pencilan terkecil.

$i_2$  = Jumlah data pencilan terbesar.

Dengan demikian winsorized MOM untuk variabel ke- $j$  dan dugaan matriks kovarians Winsorized antara variabel  $W_i$  dan  $W_j$  didefinisikan sebagai berikut [9]:

$$\bar{w}_{ij} = \frac{1}{m_j} \left[ \sum_{k=1}^{m_j} w_{ij} \right] \quad (12)$$

Dimana,

$\bar{w}_{ij}$  = Penaksir rata-rata baru untuk variabel ke- $j$

$w_{ij}$  = Observasi baru ke- $i$  untuk variabel ke- $j$

$m_j$  = Ukuran data set untuk setiap variable

### 2.6. Skala Estimator $S_n$

Dalam penelitian ini akan digunakan skala estimator  $S_n$ . Estimator ini telah diidentifikasi lebih efisien 58.23% dari  $MAD_n$ , dan memiliki *bounded influence function* (nilainya tidak menuju tak hingga jika nilai  $X$  sangat besar), yang merupakan salah satu dari sifat yang penting dalam *robust* estimator. Estimator  $S_n$  memiliki perhitungan yang mudah. Estimator  $S_n$  untuk sampel  $X$  dengan  $p$  variabel prediktor dan pada kelompok  $g$  dengan distribusi sebarang didefinisikan dengan

$$S_{n,g} = [S_{n,1} \quad S_{n,2} \quad \dots \quad S_{n,j} \quad \dots \quad S_{n,p}] \quad (13)$$

dengan  $g = 1, 2, \dots, k$ ;  $j = 1, 2, \dots, p$  dan

$$S_{n,j} = 1.1926 \text{ med}_j \{ \text{med}_i |X_{ij} - X_{hj}| \} \quad (14)$$

dengan  $h = 1, 2, \dots, n_g$ ,  $\text{med}_j$  dan  $\text{med}_i$  adalah median sampel dari pengamatan ke- $i$  dan median dari variabel prediktor  $j$  [8].

## 2.7. Evaluasi Hasil Klasifikasi

### 1. Apparent Error Rate

*Apparent Error Rate* (APER) merupakan salah satu evaluasi ketepatan klasifikasi yang bertujuan untuk menghitung peluang kesalahan klasifikasi, sehingga dapat diketahui persentase pengamatan pada populasi yang salah diklasifikasikan oleh fungsi diskriminan yang terbentuk. APER dapat dihitung dengan terlebih dahulu membuat tabel klasifikasi. Berikut klasifikasi untuk dua kelompok

Tabel 1 Tabel Klasifikasi

Kelompok Awal	Kelompok Prediksi		Jumlah Pengamatan
	1	2	
1	$n_{(1)(1)}$	$n_{(1)(2)}$	$n_{(1)}$
2	$n_{(2)(1)}$	$n_{(2)(2)}$	$n_{(2)}$

Setelah dibuat tabel klasifikasi, maka sudah dapat ditentukan APER. APER dapat didefinisikan dengan sebagai berikut.

$$APER = \frac{n_{(1)(2)} + n_{(2)(1)}}{n_{(1)} + n_{(2)}} \times 100\% \quad (15)$$

dengan

$n_{(1)(1)}$  = Jumlah pengamatan dari kelompok 1 diklasifikasikan sebagai kelompok 1

$n_{(1)(2)}$  = Jumlah pengamatan kelompok 1 namun diprediksi sebagai kelompok 2

$n_{(2)(1)}$  = Jumlah pengamatan kelompok 2 namun diprediksi sebagai kelompok 1

$n_{(2)(2)}$  = Jumlah pengamatan dari kelompok 2 diklasifikasikan sebagai kelompok 2

$n_{(1)}$  = Jumlah pengamatan pada kelompok 1

$n_{(2)}$  = Jumlah pengamatan pada kelompok 2

### 2. Press's Q Statistic

Pengujian ini dilakukan dengan membandingkan banyaknya klasifikasi yang benar dengan banyaknya total ukuran sampel dan banyaknya kelompok. Hasil yang diperoleh akan dibandingkan dengan nilai kritis (nilai *chi-square* dengan derajat kebebasan 1 pada derajat kepercayaan yang diinginkan). Hipotesis yang digunakan sebagai berikut

$$H_0 : \text{Klasifikasi konsisten}$$

$$H_1 : \text{Klasifikasi tidak konsisten}$$

Statistik uji dari *Press's Q* yang dapat didefinisikan dengan sebagai berikut:

$$Press's Q = \frac{(N - (n_u \times k))^2}{N(k - 1)} \quad (16)$$

dengan

$N$  = total sampel

$n_u$  = jumlah pengamatan yang tepat diklasifikasikan

$k$  = jumlah kelompok

Kriteria uji adalah terima  $H_0$  a jika  $Press's Q > \chi_{\alpha,1}^2$ , selain itu tolak  $H_0$

### 3. Hasil dan Diskusi

#### 3.1 Uji Homogenitas Vektor Rata-rata

Uji homogenitas vektor rata-rata dilakukan dengan menggunakan statistik uji *V-Bartlett* dengan hipotesis sebagai berikut.

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

dengan tingkat signifikansi  $\alpha = 0,05$

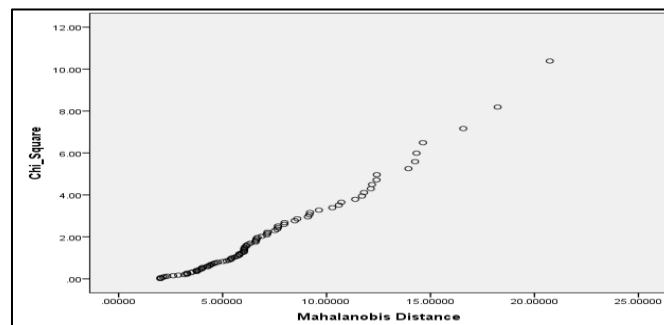
$$\Lambda = \frac{|W|}{|B + W|} = \frac{\left| \sum_{g=1}^k \sum_{i=1}^{n_g} (\mathbf{X}_{gi} - \bar{\mathbf{X}}_g)(\mathbf{X}_{gi} - \bar{\mathbf{X}}_g)^t \right|}{\left| \sum_{g=1}^k \sum_{i=1}^{n_g} (\mathbf{X}_{gi} - \bar{\mathbf{X}})(\mathbf{X}_{gi} - \bar{\mathbf{X}})^t \right|} = 0,7795$$

$$\begin{aligned} V &= - \left[ (90 - 1) - \frac{7 + 2}{2} \right] \ln(0.7795) \\ &= - \left[ 89 - \frac{9}{2} \right] (-0.2491) \\ &= -84.5 (-0.2491) \\ &= 21.04895 \end{aligned}$$

Sstatistik uji *V-Bartlett* = 21.044, karena statistik uji lebih besar dari  $\chi_{0,05,7}^2$  (14.067), maka  $H_0$  ditolak sehingga fungsi pembeda (*discriminant*) layak disusun untuk mengkaji hubungan antarkelompok dan seluruh variabel prediktor yang digunakan dianggap dapat membedakan antara kelompok prediabetes dan diabetes dengan baik.

#### 3.2 Uji Normalitas Multivariat

Pengujian normalitas multivariat dilakukan dengan menghitung jarak mahalnobis ( $d_i^2$ ) untuk setiap pengamatan dan setelah itu akan diinterpretasikan ke dalam plot kuantil *Chi-square* dengan titik koordinat  $\left( d_i^2; \chi_{\frac{1}{n}(i-0,5),p}^2 \right)$ . Hasil intepretasi pengujian normalitas multivariat pada plot *Chi-square* dengan menggunakan *software SPSS 23* adalah sebagai berikut.



Gambar 1 Plot Kuantil *Chi-square*



Dengan menggunakan *syntax Software MATLAB 2009* maka terdapat 56.66 % nilai  $d_i^2 \leq \chi_{\frac{1}{n}(i-0,5),p}^2$ . Karena titik-titik pada plot kuantil *Chi-square* menyebar mendekati garis lurus dan terdapat lebih dari 50% (56,66%) nilai  $d_i^2 \leq \chi_{\frac{1}{n}(i-0,5),p}^2$  maka dapat disimpulkan bahwa data berdistribusi normal multivariat.

### 3.3 Uji Homogenitas Matriks Variansi Kovariansi

Uji homogenitas matriks variansi kovariansi dilakukan dengan menggunakan statistik uji *Box's M* dengan hipotesis sebagai berikut.

$$H_0 : \Sigma_1 = \Sigma_2$$

$$H_1 : \Sigma_1 \neq \Sigma_2$$

dengan tingkat signifikansi  $\alpha = 0,05$ .

Berdasarkan perhitungan diperoleh nilai

$$u = \left[ \frac{1}{n_1 - 1} + \frac{1}{n_{2-1}} - \frac{1}{n_1 + n_1 - 2} \right] \left[ \frac{2p^2 + 3p - 1}{6(p + 1)(k - 1)} \right]$$

$$= \left[ \frac{1}{20 - 1} + \frac{1}{70 - 1} - \frac{1}{20 + 70 - 2} \right] \left[ \frac{2(7^2) + 3(7) - 1}{6(7 + 1)(2 - 1)} \right]$$

$$= \left[ \frac{1}{19} + \frac{1}{69} - \frac{1}{88} \right] \left[ \frac{98 + 21 - 1}{48(1)} \right]$$

$$= \left[ \frac{1}{19} + \frac{1}{69} - \frac{1}{88} \right] \left[ \frac{118}{48} \right]$$

$$= 0.1371$$

$$M = [n_1 + n_2 - 2] \ln|\mathbf{S}_{gab}| - [(n_1 - 1) \ln|\mathbf{S}_g| + (n_2 - 1) \ln|\mathbf{S}_g|]$$

$$= [(88)(34.2151) - [(19)(33.7409) + (69)(33.8018)]]$$

$$= 37.5284$$

sehingga

$$C = (1 - u)M$$

$$= (1 - 0.1371)37.5284$$

$$= 32.3841$$

Karena  $C = 32,384$  lebih kecil dari  $\chi_{0,05,28}^2 = 41,33714$ , maka  $H_0$  diterima yang berarti matriks variansi kovariansi antara kelompok prediabetes dan diabetes sama (homogen). Oleh karena itu, disebabkan asumsi homogenitas terpenuhi jadi analisis diskriminan linear dapat digunakan.

### 3.4 Pendeteksian Pencilan

Pendeteksian data pencilan dilakukan dengan menggunakan *syntax softwarew MATLAB 2009*. Pendeteksian pencilan dilakukan dengan menghitung jarak mahalanobis *robust*. Jarak mahalanobis *robust* akan dibandingkan dengan  $\chi_{0,05,7}^2 = 14.067$ . Data akan dikatakan sebagai pencilan jika  $d_{i,WMO M-S_n}^2 > \chi_{0,05,7}^2$ . Berdasarkan proses pendeteksian pencilan yang dilakukan, diperoleh hasil bahwa 8 dari 90 data pengamatan terdeteksi sebagai pengamatan pencilan. Pencilan ini terdapat paa pengamatan ke-5, ke-9, ke-12, ke-48, ke-55, ke-67, ke-83 dan ke-87.

### 3.5 Analisis Diskriminan Linear Robust WMOM- $S_n$

Dilakukan penaksiran parameter fungsi diskriminan pada analisis diskriminan linier *robust* yaitu  $\bar{X}_{WMOM-S_n}$  dan  $S_{WMOM-S_n}$  yang merupakan vektor rata-rata dan matriks variansi kovariansi dengan metode WMOM- $S_n$ . Sebelum dilakukan penaksiran maka data perlu distandarisasi terlebih dahulu. Dengan menggunakan data yang telah distandarisasi maka akan diperoleh vektor rata-rata dan matriks variansi kovariansi dengan metode WMOM- $S_n$  sebagai berikut :

Vektor rata-rata kelompok 1 (Prediabetes):

$$\bar{X}_{1,WMOM-S_n} = [0.6249 \quad -0.1618 \quad -0.5540 \quad -0.3424 \quad 0.0414 \quad 0.2858 \quad 0.4982]$$

Vektor rata-rata kelompok 2 (Diabetes) :

$$\bar{X}_{2,WMOM-S_n} = [-0.1818 \quad 0.0425 \quad 0.0948 \quad -0.0161 \quad -0.0969 \quad -0.0680 \quad -0.1913]$$

Matriks variansi kovariansi kelompok 1 (Prediabetes) :

$$S_{1,WMOM-S_n} = \begin{bmatrix} 0.9632 & -0.2550 & -0.1701 & -0.0424 & 0.0574 & 0.0245 & -0.0931 \\ -0.2550 & 0.7689 & -0.1021 & -0.1231 & -0.4176 & -0.0056 & 0.1809 \\ -0.1701 & -0.1021 & 0.9137 & 0.3018 & 0.0744 & -0.0506 & 0.1515 \\ -0.0424 & -0.1231 & 0.3018 & 0.4265 & -0.0085 & -0.0337 & -0.1037 \\ 0.0574 & -0.4176 & 0.0744 & -0.0085 & 0.8224 & 0.0821 & -0.2067 \\ 0.0245 & -0.0056 & -0.0506 & -0.0337 & 0.0821 & 0.0287 & -0.0226 \\ -0.0931 & 0.1809 & 0.1515 & -0.1037 & -0.2067 & -0.0226 & 0.8029 \end{bmatrix}$$

Matriks variansi kovariansi kelompok 2 (Diabetes) :

$$S_{2,WMOM-S_n} = \begin{bmatrix} 0.8411 & 0.1503 & -0.0180 & -0.0946 & 0.2222 & 0.2560 & 0.2253 \\ 0.1503 & 1.0501 & -0.0459 & 0.1294 & -0.0180 & 0.1502 & 0.3243 \\ -0.0180 & -0.0459 & 0.6852 & 0.1362 & 0.2204 & 0.2116 & -0.1678 \\ -0.0946 & 0.1294 & 0.1362 & 0.7389 & -0.0080 & 0.0826 & 0.0484 \\ 0.2222 & -0.0180 & 0.2204 & -0.0080 & 0.7507 & 0.5751 & -0.0896 \\ 0.2560 & 0.1502 & 0.2116 & 0.0826 & 0.5751 & 1.0380 & -0.0798 \\ 0.2253 & 0.3243 & -0.1678 & 0.0484 & -0.0896 & -0.0798 & 0.7201 \end{bmatrix}$$

Matriks variansi kovariansi gabungan

$$S_{gab,WMOM-S_n} = \begin{bmatrix} 0.8674 & 0.0628 & -0.0508 & -0.0833 & 0.1866 & 0.2060 & 0.1566 \\ 0.0628 & 0.9894 & -0.0581 & 0.0749 & -0.1043 & 0.1166 & 0.2933 \\ -0.0508 & -0.0581 & 0.7345 & 0.1720 & 0.1889 & 0.1550 & -0.0988 \\ 0.0833 & 0.0749 & 0.1720 & 0.6715 & -0.0081 & 0.0575 & 0.0155 \\ 0.1866 & -0.1043 & 0.1889 & -0.0081 & 0.7662 & 0.4687 & -0.1149 \\ 0.2060 & 0.1166 & 0.1550 & 0.0575 & 0.4687 & 0.8200 & -0.0675 \\ 0.1566 & 0.2933 & -0.0988 & 0.0155 & -0.1149 & -0.0675 & 0.7380 \end{bmatrix}$$

diperoleh fungsi diskriminan linier *robust* sebagai berikut :

$$d(Z) = -0.6352 + 0.5868Z_1 - 0.6759Z_2 - 0.8246Z_3 - 0.2128Z_4 - 0.1371Z_5 + 0.7128Z_6 + 1.0164Z_7$$

Setelah fungsi diskriminan linear *robust* telah terbentuk, maka akan diperoleh skor diskriminan untuk setiap pengamatan. Fungsi diskriminan yang digunakan merupakan fungsi diskriminan dengan data yang telah distandarisasi. Untuk memperoleh fungsi diskriminan yang setara dengan fungsi diskriminan linear robust pada data standarisasi untuk data yang tidak distandarisasi digunakan cara sebagai berikut.

$$Z_i = \frac{X_i - \mu_i}{SD_i}, i = 1,2,3,4,5,6,7$$

$$0.5868Z_1 = 0.5868 \left( \frac{X_i - \mu_i}{SD_i} \right) = \frac{0.5868X_1}{SD_1} - \frac{0.5868\mu_1}{SD_1}$$

$$= \frac{0.5868X_1}{8.811} - \frac{0.5868(53.644)}{8.811} = 0.0665 X_1 - 3.7058$$

$$-0.6759Z_2 = -0.6759 \left( \frac{X_i - \mu_i}{SD_i} \right) = \frac{-0.6759 X_2}{SD_2} - \frac{-0.6759\mu_2}{SD_2}$$

$$= \frac{-0.6759 X_2}{11.570} - \frac{-0.6759(63.8222)}{11.570} = -0.05841 X_2 + 3.7284$$

Cara ini dilakukan sampai  $Z_7$  ditransformasi dan diperoleh fungsi dsikriminan yaitu

$$d(\mathbf{X}) = -5.7762 + 0.0665X_1 - 0.0584X_2 - 0.0195X_3 - 0.0003X_4 - 0.0102X_5 \\ + 0.1018X_6 + 0.6111X_7$$

### 3.6 Hasil Pengklasifikasian Pengamatan

Terdapat 15 pengamatan yang diprediksi salah diklasifikasikan dengan rincian 9 pengamatan dari kelompok 1 (Prediabetes) namun diprediksi sebagai kelompok 2 (Diabetes) dari 20 pengamatan dan 6 pengamatan dari kelompok 2 namun diprediksi sebagi kelompok 1 dari 70 pengamatan.

### 3.7 Evaluasi Hasil

#### a. Apparent Error Rate (APER)

Sebelum menghitung APER, maka hasil klasifikasi perlu ditampilkan dalam Tabel 2.

Tabel 2. Klasifikasi Pengamatan Analisis Diskriminan Linear Robust WMOM- $S_n$

Kelompok Awal	Kelompok Prediksi		Jumlah Pengamatan
	1	2	
1	11	9	20
2	6	64	70
Jumlah Pengamatan	17	73	90

Berdasarkan Tabel 2 dengan menggunakan rumus pada persamaan (15) maka diperoleh

$$APER = \frac{9 + 6}{90} \times 100\% = 16,67\%$$

yang menunjukkan bahwa 16,67% dari pengamatan yang diprediksi salah klasifikasi.

#### **b. Press's Q Statistic**

Dengan menggunakan rumus pada persamaan (16), maka diperoleh

$$Press's Q = \frac{(90 - (75 \times 2))^2}{90(2 - 1)} = 40$$

Nilai *Press's Q* yang diperoleh akan dibandingkan dengan nilai  $\chi^2_{(1,0.05)} = 3,841$ . Karena *Press's Q*  $> \chi^2_{(1,0.05)}$  maka  $H_0$  diterima, ini berarti bahwa pengklasifikasi kelompok prediabetes dan diabetes dilakukan secara konsisten.

## **4. Kesimpulan**

Fungsi diskriminan linear *robust* dengan metode *Winsorized Modified One-Step M-Estimator* dengan skala estimator  $S_n$  pada penderita diabetes dan prediabetes di Puskesmas Kecamatan Pesanggrahan Kota Jakarta Selatan Provinsi DKI Jakarta periode Desember 2016 - Januari 2017 adalah

$$d(\mathbf{X}) = -5.7762 + 0.0665X_1 - 0.0584X_2 - 0.0195X_3 - 0.0003X_4 - 0.0102X_5 \\ + 0.1018X_6 + 0.6111X_7$$

dengan nilai APER 16,67% yang berarti presentase kesalahan klasifikasi sebesar 16,67%.

## **Daftar Pustaka**

- [1] William R. Dillon, M. G. *Multivariate Analysis Methods and Application*. New York: John Wiley&Sons.Inc. 1984.
- [2] Melik, H. N. *Robust Linear Discriminant Robust Linear Discriminant Analysis Using MOM-Qn And WMOM-Qn Estimators: Coordinate-Wise Approach*. Malaysia: Universiti Utara Malaysia. 2017.
- [3] Putri, B. S. (2017). *Pengklasifikasian Kejadian Diabetes dengan Analisis Diskriminan Kuadratik*. Jakarta: Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Jakarta.
- [4] Mattjik, A., dan Sumertajaya, M. *Sidik Peubah Ganda dengan Menggunakan SAS*. Bogor: Institut Teknologi Bandung. 2011.
- [5] Johnson, R. A., dan Wichern, D. W. *Applied Multivariate Statistical Analysis*. USA: Prentice-Hall. 2002.
- [6] Harlow, L. *The Essence of Multivariate Thinking: Basic Themes and Methods*. Routledge. 2014.

- [7] Wilcox, R. Multiple comparisons among dependen groups based on a modified one-step M-Estimator. *Journal of Applied Statistics*, 30(10), 1231-1241. 2002.
- [8] Yahaya, S. S., Lim, Y. F., dan Ali, H. Robust Linear Discriminant Analysis with Highest Breakdown Point Estimator. *Journal of Telecommunication, Electronic and Computer Engineering*, Vol 10, 1-11. 2013.