BAB I
PENDAHULUAN

1.1 Latar Belakang

Analisis diskriminan adalah analisis yang digunakan untuk memodelkan hubungan antara satu peubah respon yang memiliki skala non metrik (nominal atau ordinal) dengan lebih dari satu peubah prediktor yang memiliki skala metrik (interval atau rasio). Analisis ini bertujuan untuk mengklasifikasikan obyek ke dalam kelompok yang telah diketahui dari informasi awal dan menghitung risiko dari kemungkinan kesalahan pengklasifikasian. Menurut Hair, dkk (1998), asumsi yang harus dipenuhi pada analisis diskriminan yaitu peubah prediktor menyebar normal multivariat dan matriks ragam peragam yang homogen untuk setiap kelompok. Selain itu, identifikasi pencilan pada data harus dilakukan karena pencilan dapat menyebabkan asumsi analisis diskriminan menjadi tidak terpenuhi dan klasifikasi obyek yang dihasilkan menjadi kurang tepat. Pencilan adalah data yang letaknya jauh dari pola data pada umumnya. Analisis diskriminan kurang tepat digunakan ketika terdapat pencilan pada data, sehingga analisis yang tepat digunakan menurut Rousseuw,dkk (1998) yaitu analisis diskriminan linier *robust*. Menurut Croux, dkk (2008), analisis tersebut merupakan perkembangan dari analisis diskriminan klasik. Konsep dasar dari analisis diskriminan linier *robust* yaitu mengganti vektor rata-rata dan matriks ragam peragam yang diperoleh dari perhitungan menggunakan data awal dengan vektor rata-rata dan matriks ragam peragam yang *robust* dalam pendugaan parameter menggunakan penduga-S dengan algoritma *SURREAL*.

Berdasarkan kriteria WHO pada tahun 1961, berat badan bayi saat dilahirkan dibagi menjadi dua kelompok yaitu Berat Bayi Lahir Rendah (BBLR) dan berat bayi lahir normal. BBLR yaitu bayi lahir dengan berat badan kurang dari atau sama dengan 2500 gram, sedangkan berat bayi lahir normal yaitu bayi lahir dengan berat badan lebih dari 2500 gram. Berat bayi saat dilahirkan dipengaruhi oleh keadaan janin itu sendiri, selain itu kondisi ibu secara langsung juga mempengaruhi berat bayi saat dilahirkan. Tingginya angka BBLR dapat mempengaruhi kualitas sumber daya manusia di masa depan. Oleh karena itu, berbagai upaya perlu dilakukan untuk menurunkan angka BBLR. Salah satu cara yaitu melakukan klasifikasi ulang obyek (bayi baru lahir) ke dalam kelompok berat bayi baru lahir berdasarkan peubah-peubah yang mewakili keadaan ibu saat hamil. Klasifikasi ulang tersebut dapat dilakukan dengan analisis diskriminan. Namun karena terdapat pencilan pada data, maka analisis yang tepat digunakan adalah analisis diskriminan linier *robust*.

Penelitian terdahulu terkait dengan analisis diskriminan yaitu penelitian Hayati (2009). Peubah yang digunakan merupakan peubah-peubah yang mempengaruhi status gizi buruk secara tidak langsung di suatu kabupaten/kota di Jawa Timur. Sebanyak 38 kabupaten/kota di Jawa Timur dikelompokkan ke dalam tiga kelompok status gizi masyarakat yaitu 0 = normal, 1 = gizi kurang, dan 2 = gizi buruk. Namun, adanya pencilan pada data tidak diperhatikan. Apabila pencilan tersebut diatasi, maka akan diperoleh ketepatan klasifikasi yang tinggi.

Pada penelitian ini, data yang digunakan yaitu data berat bayi baru lahir dan peubah-peubah yang mewakili keadaan ibu saat hamil. Karena penelitian ini bertujuan untuk melakukan klasifikasi ulang bayi baru lahir, maka digunakan analisis diskriminan. Sebelum analisis diskriminan diterapkan, hasil identifikasi pada data menunjukkan adanya pencilan. Adanya pencilan ini menyebabkan beberapa asumsi (peubah prediktor menyebar normal multivariat dan kehomogenan matriks ragam peragam) pada analisis diskriminan tidak terpenuhi. Oleh sebab itu analisis diskriminan klasik kurang tepat digunakan, sebagai penggantinya harus digunakan analisis diskriminan linier *robust*. Dari hasil analisis tersebut dengan menggunakan algoritma *SURREAL* pada penduga-S, maka akan diperoleh penduga *robust* (penduga yang kekar terhadap pencilan) yang digunakan untuk membentuk fungsi diskriminan linier *robust.* Fungsi tersebut digunakan untuk klasifikasi ulang obyek, sehingga diperoleh klasifikasi yang baru. Setelah itu, dilakukan evaluasi fungsi klasifikasi untuk mengetahui seberapa besar ketepatan klasifikasi dari analisis diskriminan linier *robust*.

Hal-hal tersebut di atas menjadi motivasi pada penelitian ini untuk melakukan klasifikasi ulang bayi baru lahir berdasarkan peubah-peubah yang mewakili keadaan ibu saat hamil dengan menggunakan analisis diskriminan linier *robust*. Dari analisis tersebut akan diperoleh fungsi diskriminan linier *robust* yang digunakan untuk mendapatkan klasifikasi yang baru. Hasil dari klasifikasi tersebut diharapkan dapat menjadi bahan pertimbangan untuk mencegah kelahiran BBLR berdasarkan peubah-peubah yang mewakili keadaan ibu saat hamil serta dapat dilakukan usaha perbaikan untuk menurunkan jumlah kelahiran BBLR.

1.2 Rumusan Masalah

Permasalahan yang ingin dibahas pada penelitian ini adalah:

1. Bagaimana pengaruh *breakdown point* yang sudah ditetapkan pada algoritma *SURREAL* terhadap ketepatan klasifikasi dari analisis diskriminan linier *robust*?
2. Bagaimana fungsi yang terbentuk dari analisis diskriminan linier *robust* khususnya pada pengklasifikasian bayi baru lahir?
3. Bagaimana prediksi dan ketepatan klasifikasi ulang bayi baru lahir berdasarkan peubah-peubah yang mewakili keadaan ibu saat hamil?

1.3 Tujuan Penelitian

Tujuan yang ingin dicapai dalam penelitian ini adalah:

1. Menganalisis seberapa besar pengaruh *breakdown point* yang sudah ditetapkan pada algoritma *SURREAL* terhadap ketepatan klasifikasi dari analisis diskriminan linier *robust*.
2. Menduga parameter *robust* untuk membentuk fungsi diskriminan linier *robust* yang nantinya akan digunakan untuk pengklasifikasian ulang bayi baru lahir.
3. Melakukan prediksi dan menghitung ketepatan dari klasifikasi ulang bayi baru lahir berdasarkan peubah-peubah yang mewakili keadaan ibu saat hamil.

1.4 Manfaat Penelitian

Manfaat yang diperoleh dari penelitian ini adalah :

1. Mengetahui *breakdown point* yang tepat digunakan pada kasus pengklasifikasian ulang bayi baru lahir berdasarkan algoritma *SURREAL* agar diperoleh ketepatan klasifikasi yang tinggi.

2. Mengetahui fungsi diskriminan linier *robust* dari data yang mengandung pencilan, sehingga fungsi tersebut dapat digunakan untuk pengklasifikasian ulang bayi baru lahir.
3. Mengetahui prediksi dan ketepatan klasifikasi ulang bayi baru lahir. Berdasarkan hasil prediksi tersebut, diharapkan dapat digunakan sebagai bahan pertimbangan untuk mencegah kelahiran BBLR berdasarkan peubah-peubah yang mewakili keadaan ibu saat hamil serta dapat dilakukan usaha perbaikan untuk menurunkan jumlah kelahiran BBLR.

1.5 Batasan Masalah

Penelitian ini dibatasi pada masalah:

1. Analisis yang digunakan adalah analisis diskriminan linier *robust* untuk dua kelompok. Selain itu, metode pendugaan parameter yang digunakan yaitu metode penduga-S dengan algoritma *SURREAL*.
2. Data yang digunakan adalah berat bayi baru lahir mulai bulan Januari 2012 sampai bulan November 2012 yang dilahirkan secara normal dan persalinan dengan tindakan. Selain itu, penelitian ini juga menggunakan data mengenai peubah-peubah yang mewakili keadaan ibu saat hamil yaitu usia ibu, jumlah anak, berat badan ibu prahamil, ukuran Lingkar Lengan Atas (LILA) ibu hamil, kadar Hb ibu hamil dan jarak kelahiran. Data tersebut diperoleh dari buku Kohort Bidan Desa di Puskesmas Manyar Kabupaten Gresik.

BAB II
TINJAUAN PUSTAKA

2.1 Analisis Diskriminan

Analisis diskriminan tepat digunakan untuk memodelkan hubungan antara satu peubah respon yang memiliki skala non metrik atau berupa kategori (nominal atau ordinal) dengan lebih dari satu peubah prediktor yang memiliki skala metrik atau kuantitatif (interval atau rasio) (Hair, dkk., 1998).

Menurut Sharma (1996), analisis diskriminan merupakan salah satu metode dependensi yaitu adanya peubah dependen dan peubah independen dalam analisis multivariat yang dapat digunakan untuk mengetahui ketepatan klasifikasi obyek. Analisis diskriminan bertujuan untuk mendapatkan suatu fungsi diskriminan yang dapat digunakan untuk memisahkan obyek sesuai dengan kelompoknya, selain itu dapat juga digunakan untuk memprediksi kelompok dari suatu obyek baru yang diamati.

Analisis diskriminan digunakan ketika semua kelompok telah diketahui dari informasi awal atau informasi prior dengan tujuan untuk mengelompokkan atau mengklasifikasikan obyek ke dalam kelompok yang telah diketahui tersebut. Analisis diskriminan hanya mengelompokkan setiap obyek ke dalam satu dari beberapa kelompok dan analisis ini harus melakukan perhitungan untuk mengetahui seberapa besar risiko dari kemungkinan kesalahan pengklasifikasian (Hardle dan Simar, 2003).

Menurut Hair, dkk (1998), asumsi untuk analisis diskriminan berhubungan dengan proses pendugaan dan prosedur pengklasifikasian yang memberikan pengaruh dalam interpretasi hasil. Asumsi yang harus dipenuhi dalam analisis diskriminan yaitu peubah prediktor menyebar normal multivariat dan matriks ragam peragam yang homogen untuk setiap kelompok. Jika kedua asumsi tersebut tidak terpenuhi maka akan berdampak pada ketepatan pengklasifikasian dan kesimpulan yang diperoleh.

2.2 Asumsi yang Melandasi Analisis Diskriminan

2.2.1 Sebaran Normal Multivariat

Menurut Johnson dan Wichern (2002), analisis multivariat dapat digunakan apabila data memenuhi asumsi sebaran normal multivariat. Analisis multivariat merupakan suatu analisis yang melibatkan lebih dari dua peubah sedangkan sebaran normal multivariat merupakan suatu bentuk sebaran di mana masing-masing peubah dari data tersebut memenuhi sifat normalitas.

Analisis diskriminan memiliki beberapa asumsi yang harus terpenuhi salah satunya yaitu peubah prediktor menyebar normal multivariat. Vektor $\underline{\boldsymbol{X}}^T = (X_1, X_2, \dots, X_p)$ dikatakan menyebar normal multivariat dengan parameter $\underline{\boldsymbol{\mu}}$ dan $\boldsymbol{\Sigma}$ jika mempunyai fungsi kepekatan peluang :

$$f(\underline{\boldsymbol{x}}) = \frac{1}{(\sqrt{2\pi})^p \,|\boldsymbol{\Sigma}|^{1/2}} e^{-\frac{1}{2}(\underline{\boldsymbol{x}}-\underline{\boldsymbol{\mu}})^T \boldsymbol{\Sigma}^{-1} (\underline{\boldsymbol{x}}-\underline{\boldsymbol{\mu}})} \quad (2.1)$$

Jika $\underline{\boldsymbol{X}}$ menyebar normal multivariat dengan fungsi kepekatan peluang seperti persamaan (2.1), maka dapat dikatakan bahwa $\left(\underline{\boldsymbol{x}} - \underline{\boldsymbol{\mu}}\right)^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} \left(\underline{\boldsymbol{x}} - \underline{\boldsymbol{\mu}}\right)$ menyebar $\chi^2_{p,\alpha}$, di mana p adalah banyak peubah prediktor. Hal ini dapat dibuktikan sebagai berikut:

$$\sum_{i=1}^{n} Z_i^2 = \sum_{i=1}^{n} \left[\frac{(x_i-\mu)}{\sigma}\right]^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{(n-1)s^2}{\sigma^2} \quad (2.2)$$

di mana :

$$s^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{(n-1)} \quad (2.3)$$

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} \quad (2.4)$$

Pada Sebaran Normal Multivariat:

$$\left[\frac{\left(\underline{\boldsymbol{x}}-\underline{\boldsymbol{\mu}}\right)}{\sigma}\right]^2 = \left(\underline{\boldsymbol{x}} - \underline{\boldsymbol{\mu}}\right)^T (\sigma^2)^{-1} \left(\underline{\boldsymbol{x}} - \underline{\boldsymbol{\mu}}\right) = \left(\underline{\boldsymbol{x}} - \underline{\boldsymbol{\mu}}\right)^T (\boldsymbol{\Sigma})^{-1} \left(\underline{\boldsymbol{x}} - \underline{\boldsymbol{\mu}}\right) \quad (2.5)$$

Jadi, dapat dikatakan bahwa:

$\sum_{i=1}^{n} Z_i^2 \sim \chi^2$ yaitu $\left(\underline{\boldsymbol{x}} - \underline{\boldsymbol{\mu}}\right)^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} \left(\underline{\boldsymbol{x}} - \underline{\boldsymbol{\mu}}\right)$ menyebar $\chi^2_{p,\alpha}$

Pengujian sebaran normal multivariat dapat dilakukan dengan cara membuat Q-Q plot yaitu plot antara jarak *Mahalanobis* (d_i^2) dengan tabel $\chi^2_{p,\alpha}$.

Adapun langkah-langkah untuk membuat Q-Q plot adalah sebagai berikut:

1. Menentukan nilai vektor rata-rata $\underline{\bar{\boldsymbol{x}}}$ berukuran *p* x 1:
$$\underline{\bar{\boldsymbol{x}}}_{(p\text{x}1)} = \frac{1}{n}\sum_{i=1}^{n} \underline{\boldsymbol{x}}_i \tag{2.6}$$
di mana *n* adalah banyak pengamatan.
2. Menentukan matriks ragam peragam $\boldsymbol{S}$ berukuran *p* x *p*:
$$\boldsymbol{S}_{(p\text{x}p)} = \frac{1}{n-1}\sum_{i=1}^{n}\left(\underline{\boldsymbol{x}}_i - \underline{\bar{\boldsymbol{x}}}\right)\left(\underline{\boldsymbol{x}}_i - \underline{\bar{\boldsymbol{x}}}\right)^T \tag{2.7}$$
di mana $\underline{\boldsymbol{x}}_i$ adalah vektor pengamatan ke-*i*.
3. Menghitung jarak *Mahalanobis* (d_i^2) untuk setiap pengamatan dengan vektor rata-ratanya:
$$d_i^2 = \left(\underline{\boldsymbol{x}}_i - \underline{\bar{\boldsymbol{x}}}\right)^T \boldsymbol{S}^{-1}\left(\underline{\boldsymbol{x}}_i - \underline{\bar{\boldsymbol{x}}}\right)\ ,\ i = 1, 2, \dots, n \tag{2.8}$$
4. Mengurutkan jarak *Mahalanobis* (d_i^2) setiap pengamatan dari nilai terkecil sampai nilai terbesar:
$$d_{(1)}^2 \le d_{(2)}^2 \le d_{(3)}^2 \le \dots \le d_{(n)}^2 \tag{2.9}$$
5. Menentukan nilai $q_i = \chi^2_{p,\alpha}$, di mana *p* adalah banyak peubah prediktor, $\alpha = (i\text{-}0.5)/n$ dan nilai $\chi^2_{p,\alpha}$ diperoleh dari tabel χ^2.
6. Membuat plot antara d_i^2 dengan q_i

Hipotesis yang digunakan untuk pengujian sebaran normal multivariat adalah:

H_0 : Data menyebar normal multivariat

H_1 : Data tidak menyebar normal multivariat

Menurut Sharma (1996), H_0 diterima yaitu data menyebar normal multivariat apabila Q-Q plot dapat didekati dengan garis lurus (linier). Pengujian kenormalan multivariat dapat dilakukan dengan menghitung dua macam ukuran statistik yaitu *skewness* ($b_{1,p}$) dan *kurtosis* ($b_{2,p}$) sebagai berikut (Matjik dan Sumertajaya, 2011):

$$b_{1,p} = \frac{1}{n^2}\sum_{i=1}^{n}\left[\left(\underline{\boldsymbol{x}}_i - \underline{\bar{\boldsymbol{x}}}\right)^T \boldsymbol{S}^{-1}\left(\underline{\boldsymbol{x}}_i - \underline{\bar{\boldsymbol{x}}}\right)\right]^3 \tag{2.10}$$

$$b_{2,p} = \frac{1}{n^2}\sum_{i=1}^{n}\left[\left(\underline{\boldsymbol{x}}_i - \underline{\bar{\boldsymbol{x}}}\right)^T \boldsymbol{S}^{-1}\left(\underline{\boldsymbol{x}}_i - \underline{\bar{\boldsymbol{x}}}\right)\right]^2 \tag{2.11}$$

Apabila $b_{1,p}\left(\frac{n}{6}\right) \leq \chi^2_{p(p+1)(p+2)/6}$ dan $\frac{b_{2,p}-p(p+2)}{\sqrt{(8p(p+2)/n)}} \leq Z_\alpha$, maka dapat diputuskan untuk menerima H_0 yaitu data menyebar normal multivariat.

2.2.2 Kehomogenan Matriks Ragam Peragam

Asumsi analisis diskriminan yang harus dipenuhi selain peubah prediktor menyebar normal multivariat yaitu kehomogenan matriks ragam peragam. Hipotesis yang digunakan untuk pengujian kehomogenan matriks ragam peragam adalah:

H_0 : $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \ldots = \boldsymbol{\Sigma}_g = \boldsymbol{\Sigma}$

H_1 : sedikitnya ada satu $\boldsymbol{\Sigma}_k$ yang berbeda, di mana k = 1,2,…,g

Misalkan $\boldsymbol{S}_k$ merupakan penduga tak bias dari $\boldsymbol{\Sigma}_k$ yaitu matriks ragam peragam untuk kelompok ke-k dengan derajat bebas $(n_k - 1)$, yaitu:

$$\boldsymbol{S}_k = \frac{1}{n_k-1}\sum_{i=1}^{n_k}\left(\underline{\boldsymbol{x}}_i - \underline{\bar{\boldsymbol{x}}}\right)\left(\underline{\boldsymbol{x}}_i - \underline{\bar{\boldsymbol{x}}}\right)^T,\ k = 1, 2, \ldots, g \tag{2.12}$$

maka diperoleh matriks ragam peragam gabungan:

$$\boldsymbol{S} = \frac{1}{\sum_{k=1}^{g} n_k - 1}\sum_{k=1}^{g}(n_k - 1)\boldsymbol{S}_k \tag{2.13}$$

di mana:

n_k adalah banyak pengamatan pada kelompok ke-k.

Pengujian asumsi kehomogenan matriks ragam peragam dapat dilakukan dengan menggunakan statistik uji Box's M sebagai berikut:

$$\text{Box's M} = MC^{-1} \tag{2.14}$$

di mana:

$$M = \sum_{k=1}^{g}(n_k - 1)\ ln\left|\boldsymbol{S}\right| - \sum_{k=1}^{g}(n_k - 1)\ ln\left|\boldsymbol{S_k}\right| \tag{2.15}$$

$$C^{-1} = 1 - \left[\frac{2p^2+3p-1}{6(p+1)(g-1)}\right]\left[\sum_{k=1}^{g}\frac{1}{(n_k-1)} - \frac{1}{\sum_{k=1}^{g}(n_k-1)}\right] \tag{2.16}$$

H_0 ditolak apabila statistik uji Box's M > $\chi^2_{\alpha,v}$, di mana v = ½ (g-1) p (p+1). Hal ini berarti bahwa matriks ragam peragam antar kelompok berbeda atau dengan kata lain bahwa asumsi kehomogenan matriks ragam peragam tidak terpenuhi. Namun jika H_0 diterima maka dapat dikatakan bahwa matriks ragam peragam untuk setiap kelompok homogen, sehingga $\boldsymbol{S}$ adalah penduga tak bias bagi matriks ragam peragam $\boldsymbol{\Sigma}$ (Timm, 1975).

2.3 Perbedaan Vektor Rata-Rata

Pada analisis diskriminan perlu diketahui apakah vektor rata-rata semua peubah prediktor dalam setiap kelompok bernilai sama atau berbeda. Jika diperoleh hasil bahwa vektor rata-rata antar kelompok sama, maka fungsi diskriminan yang nantinya terbentuk menjadi tidak signifikan. Hipotesis untuk pengujian vektor rata-rata antar kelompok adalah:

H_0 : $\underline{\boldsymbol{\mu}}_{\boldsymbol{1}} = \underline{\boldsymbol{\mu}}_{\boldsymbol{2}} = \ldots = \underline{\boldsymbol{\mu}}_{\boldsymbol{g}} = \underline{\boldsymbol{\mu}}$

H_1 : sedikitnya ada satu $\underline{\boldsymbol{\mu}}_{\boldsymbol{k}}$ yang berbeda, di mana k = 1,2,…,g

Pengujian perbedaan vektor rata-rata antar kelompok dapat dilakukan dengan menggunakan statistik uji V-Bartlett yang mengikuti sebaran Chi-Kuadrat (χ^2) dengan derajat bebas p(k-1). Apabila H_0 benar dan ukuran contoh besar, maka statistik uji V-Bartlett adalah:

$$V = -\left(n - 1 - \frac{(p+g)}{2}\right) ln\ \Lambda^* \qquad (2.17)$$

di mana:

n = banyak pengamatan
p = banyak peubah prediktor
k = banyak kelompok
Λ^* = Wilk's lambda

$$\Lambda^* = \frac{|\boldsymbol{JK_D}|}{|\boldsymbol{JK_T}|} = \frac{|\sum_{i=1}^{n_k}\sum_{k=1}^{g}(\underline{\boldsymbol{x}}_{ik} - \underline{\bar{\boldsymbol{x}}}_k)(\underline{\boldsymbol{x}}_{ik} - \underline{\bar{\boldsymbol{x}}}_k)^T|}{|\sum_{i=1}^{n_k}\sum_{k=1}^{g}(\underline{\boldsymbol{x}}_{ik} - \underline{\bar{\boldsymbol{x}}})(\underline{\boldsymbol{x}}_{ik} - \underline{\bar{\boldsymbol{x}}})^T|} \qquad (2.18)$$

dalam hal ini:

$$\boldsymbol{JK_A} = \sum_{k=1}^{g} n_k(\underline{\bar{\boldsymbol{x}}}_k - \underline{\bar{\boldsymbol{x}}})(\underline{\bar{\boldsymbol{x}}}_k - \underline{\bar{\boldsymbol{x}}})^T \tag{2.19}$$

$$\boldsymbol{JK_D} = \sum_{i=1}^{n_k} \sum_{k=1}^{g} (\underline{\boldsymbol{x}}_{ik} - \underline{\bar{\boldsymbol{x}}}_k)(\underline{\boldsymbol{x}}_{ik} - \underline{\bar{\boldsymbol{x}}}_k)^T \tag{2.20}$$

$$\boldsymbol{JK_T} = \boldsymbol{JK_A} + \boldsymbol{JK_D} = \sum_{i=1}^{n_k} \sum_{k=1}^{g} (\underline{\boldsymbol{x}}_{ik} - \underline{\bar{\boldsymbol{x}}})(\underline{\boldsymbol{x}}_{ik} - \underline{\bar{\boldsymbol{x}}})^T \tag{2.21}$$

Keterangan:
$\underline{\boldsymbol{x}}_{ik}$: vektor pengamatan ke-*i* kelompok ke-*k*.
$\underline{\bar{\boldsymbol{x}}}_k$: vektor rata-rata kelompok ke-*k*.
$\underline{\bar{\boldsymbol{x}}}$: vektor rata-rata keseluruhan.

H_0 ditolak pada taraf nyata α apabila statistik uji V-Bartlett > $\chi^2_{\alpha,p(g-1)}$, dengan kata lain yaitu terdapat perbedaan vektor rata-rata antar kelompok. Hal ini berarti bahwa fungsi diskriminan dapat digunakan untuk mengklasifikasikan obyek baru ke dalam salah satu kelompok yang telah diketahui (Johnson dan Wichern, 2002).

2.4 Pencilan

Pencilan adalah pengamatan yang dapat diidentifikasi secara jelas yang berbeda dari pengamatan yang lain. Meskipun berbeda dengan pengamatan yang lain, namun pencilan dapat menunjukkan karakteristik dari populasi. Sebaliknya, apabila pencilan bermasalah tidak mewakili populasi dan bertentangan dengan tujuan analisis maka secara serius dapat memberikan hasil uji statistik yang berbeda. Pemeriksaan atau identifikasi pencilan pada data harus dilakukan karena pencilan memberikan pengaruh pada ragam dan setelah pencilan teridentifikasi maka dapat diputuskan untuk tetap mempertahankan atau menghapus pencilan tersebut (Hair, dkk., 1998).

Menurut Rencher (2002), suatu pendekatan untuk mengatasi pencilan pada data multivariat yaitu melakukan pendugaan dengan metode *robust*. Metode ini meminimumkan pengaruh pencilan dalam pendugaan atau ketepatan model. Meskipun demikian, terkadang pencilan merupakan informasi yang penting dan sangat bermanfaat.

Identifikasi pencilan pada analisis multivariat yang melibatkan lebih dari dua peubah dibutuhkan suatu alat untuk mengukur posisi multidimensi setiap pengamatan secara relatif terhadap beberapa titik. Salah satu cara pendeteksian pencilan yaitu dengan menggunakan jarak *Mahalanobis* (d_i^2) di mana jarak ini mengukur jarak multidimensi pada setiap pengamatan dari pusat rata-rata pengamatan (Hair, dkk., 1998). Menurut Roussew dan Zomeren (1990), pengamatan dikatakan sebagai pencilan apabila $d_i > \sqrt{\chi^2_{p,0.975}}$ atau $d_i^2 > \chi^2_{p,0.975}$ dengan p adalah banyak peubah prediktor.

2.5 Pendugaan Parameter Dalam Sebaran Normal Multivariat

Ketika asumsi sebaran normal multivariat telah terpenuhi untuk suatu populasi tertentu, maka penduga parameter seringkali dapat ditentukan dengan metode MLE (*Maximum Likelihood Estimation*). Konsep dari metode ini sederhana yaitu misal diketahui vektor berukuran p x1 yaitu $\underline{\boldsymbol{x}}_1, \underline{\boldsymbol{x}}_2, \ldots, \underline{\boldsymbol{x}}_n$ yang merupakan contoh acak dari sebuah populasi yang menyebar normal multivariat dengan vektor rata-rata $\underline{\boldsymbol{\mu}}$ dan matriks ragam peragam $\boldsymbol{\Sigma}$. Kedua parameter tersebut ditentukan sedemikian rupa sehingga memaksimumkan kepekatan bersama dari pengamatan $\underline{\boldsymbol{x}}_{\boldsymbol{i}}$ yang disebut sebagai fungsi *likelihood*. Karena $\underline{\boldsymbol{x}}_{\boldsymbol{i}}$ dinyatakan sebagai suatu peubah acak, maka $\underline{\boldsymbol{x}}_{\boldsymbol{i}}$ akan saling bebas dan kepekatan bersamanya merupakan perkalian dari kepekatan $\underline{\boldsymbol{x}}_{\boldsymbol{i}}$. Sehingga fungsi *likelihood* adalah:

$$
\begin{aligned}
L\left(\underline{\boldsymbol{x}}_1, \underline{\boldsymbol{x}}_2, \ldots, \underline{\boldsymbol{x}}_n, \underline{\boldsymbol{\mu}}, \boldsymbol{\Sigma}\right) &= \prod_{i=1}^{n} f\left(\underline{\boldsymbol{x}}_{\boldsymbol{i}}, \underline{\boldsymbol{\mu}}, \boldsymbol{\Sigma}\right) \\
&= \prod_{i=1}^{n} \frac{1}{(\sqrt{2\pi})^p |\boldsymbol{\Sigma}|^{1/2}} e^{-\frac{1}{2}\left(\underline{\boldsymbol{x}}_{\boldsymbol{i}}-\underline{\boldsymbol{\mu}}\right)^T \boldsymbol{\Sigma}^{-1}(\underline{\boldsymbol{x}}_{\boldsymbol{i}}-\underline{\boldsymbol{\mu}})} \\
&= \frac{1}{(\sqrt{2\pi})^{np}\, |\boldsymbol{\Sigma}|^{n/2}} e^{-\frac{1}{2}\sum_{i=1}^{n}\left(\underline{\boldsymbol{x}}_{\boldsymbol{i}}-\underline{\boldsymbol{\mu}}\right)^T \boldsymbol{\Sigma}^{-1}(\underline{\boldsymbol{x}}_{\boldsymbol{i}}-\underline{\boldsymbol{\mu}})} \qquad (2.22)
\end{aligned}
$$

Untuk melihat bahwa $\underline{\hat{\boldsymbol{\mu}}} = \underline{\bar{\boldsymbol{x}}}$ memaksimumkan fungsi *likelihood*, maka dengan menambah dan mengurangi $\underline{\bar{\boldsymbol{x}}}$ ke dalam pangkat eksponen dalam persamaan (2.22) akan diperoleh:

$$L = \frac{1}{(\sqrt{2\pi})^{np}\,|\boldsymbol{\Sigma}|^{n/2}}\, e^{-\frac{1}{2}\sum_{i=1}^{n}\left(\underline{\boldsymbol{x}_i}-\underline{\bar{\boldsymbol{x}}}+\underline{\bar{\boldsymbol{x}}}-\underline{\boldsymbol{\mu}}\right)^T \boldsymbol{\Sigma}^{-1}(\underline{\boldsymbol{x}_i}-\underline{\bar{\boldsymbol{x}}}+\underline{\bar{\boldsymbol{x}}}-\underline{\boldsymbol{\mu}})} \qquad (2.23)$$

Ketika bentuk di atas diperluas dalam bentuk $\underline{\boldsymbol{x}_i} - \underline{\bar{\boldsymbol{x}}}$ dan $\underline{\bar{\boldsymbol{x}}} - \underline{\boldsymbol{\mu}}$ maka akan diperoleh hasil $\sum_{i=1}^{n} \underline{\boldsymbol{x}_i} - \underline{\bar{\boldsymbol{x}}}$ = 0, sehingga persamaan (2.23) menjadi:

$$L = \frac{1}{(\sqrt{2\pi})^{np}\,|\boldsymbol{\Sigma}|^{n/2}}\, e^{-\sum_{i=1}^{n}(\underline{\boldsymbol{x}_i}-\underline{\bar{\boldsymbol{x}}})^T \boldsymbol{\Sigma}^{-1}(\underline{\boldsymbol{x}_i}-\underline{\bar{\boldsymbol{x}}})/2 - n\left(\underline{\bar{\boldsymbol{x}}}-\underline{\boldsymbol{\mu}}\right)^T \boldsymbol{\Sigma}^{-1}(\underline{\bar{\boldsymbol{x}}}-\underline{\boldsymbol{\mu}})/2} \qquad (2.24)$$

Karena $\boldsymbol{\Sigma}^{-1}$ merupakan matriks yang definit positif, maka diperoleh:

$$-n\left(\underline{\bar{\boldsymbol{x}}} - \underline{\boldsymbol{\mu}}\right)^T \boldsymbol{\Sigma}^{-1}(\underline{\bar{\boldsymbol{x}}} - \underline{\boldsymbol{\mu}})/2 \leq 0 \text{ dan } 0 < e^{-n\left(\underline{\bar{\boldsymbol{x}}}-\underline{\boldsymbol{\mu}}\right)^T \boldsymbol{\Sigma}^{-1}(\underline{\bar{\boldsymbol{x}}}-\underline{\boldsymbol{\mu}})/2} \leq 1 \qquad (2.25)$$

di mana nilai L akan maksimum pada saat pangkat dari eksponen sama dengan 0. Sehingga L akan maksimum pada saat $\underline{\hat{\boldsymbol{\mu}}} = \underline{\bar{\boldsymbol{x}}}$.

Penduga *maximum likelihood* dari $\underline{\boldsymbol{\mu}}$ dan $\boldsymbol{\Sigma}$ untuk peubah acak yang menyebar normal multivariat adalah:

$$\underline{\hat{\boldsymbol{\mu}}} = \underline{\bar{\boldsymbol{x}}} = \frac{1}{n}\sum_{i=1}^{n} \underline{\boldsymbol{x}}_i \qquad (2.26)$$

dan

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n}(\underline{\boldsymbol{x}}_i - \underline{\bar{\boldsymbol{x}}})(\underline{\boldsymbol{x}}_i - \underline{\bar{\boldsymbol{x}}})^T \qquad (2.27)$$

Persamaan (2.27) merupakan penduga bias bagi matriks ragam peragam $\hat{\boldsymbol{\Sigma}}$, sehingga digunakan $\boldsymbol{S}$ sebagai penduga tak bias bagi matriks ragam peragam $\hat{\boldsymbol{\Sigma}}$ sebagai berikut:

$$\boldsymbol{S} = \frac{1}{n-1}\sum_{i=1}^{n}(\underline{\boldsymbol{x}}_i - \underline{\bar{\boldsymbol{x}}})(\underline{\boldsymbol{x}}_i - \underline{\bar{\boldsymbol{x}}})^T \qquad (2.28)$$

(Johnson dan Wichern, 2002).

2.6 Fungsi Diskriminan Linier Fisher dan Aturan Pengklasifikasian

Fungsi diskriminan pertama kali diperkenalkan oleh Ronald A. Fisher dengan menggunakan beberapa kombinasi linier dari pengamatan yang cukup mewakili populasi. Tujuan utama analisis diskriminan Fisher adalah memisahkan populasi. Selain itu, dapat juga digunakan untuk pengklasifikasian obyek. Untuk mencari kombinasi linier dari *p* peubah prediktor dapat dilakukan dengan pemilihan koefisien-koefisien yang mempunyai hasil bagi maksimum antara ragam antar kelompok dan ragam dalam kelompok.

Analisis diskriminan Fisher mencoba menghasilkan kombinasi linier terbaik antara dua atau lebih peubah prediktor yang dapat membedakan kelompok secara maksimum dan memasukkan obyek tersebut dalam suatu kelompok. Banyak kelompok (*k*) harus memenuhi $2k < p$ dengan *p* adalah banyak peubah prediktor. Fungsi diskriminan yang terbentuk tergantung dari banyak kelompok yang diidentifikasi yaitu sebanyak *k*-1 (Mattjik dan Sumertajaya, 2011).

Analisis diskriminan linier Fisher untuk dua kelompok dimulai dengan asumsi bahwa pengamatan diambil dari populasi yang menyebar normal multivariat dengan *p* dimensi yang memiliki rata-rata $\underline{\boldsymbol{\mu}_1}$ dan $\underline{\boldsymbol{\mu}_2}$. Pengelompokkan dibentuk menjadi populasi kelompok pertama (π_1) dan kelompok kedua (π_2) dengan probabilitas prior dari kelompok 1 = p_1 dan probabilitas prior dari kelompok 2 = p_2. Fisher mengasumsikan probabilitas prior dari kedua kelompok tersebut adalah sama (Johnson dan Wichern, 2002).

$\underline{\boldsymbol{X}}$ adalah suatu individu yang menggambarkan vektor acak pembeda dengan $E(\underline{\boldsymbol{x}_1}) = \underline{\boldsymbol{\mu}_1}$ dan $E(\underline{\boldsymbol{x}_2}) = \underline{\boldsymbol{\mu}_2}$ serta $\mathrm{Cov}(\underline{\boldsymbol{x}_1}) = \mathrm{Cov}(\underline{\boldsymbol{x}_2}) = \boldsymbol{\Sigma}$, dalam hal ini:

$$\underline{\boldsymbol{\mu}_k} = \begin{bmatrix} \underline{\boldsymbol{\mu}_1} \\ \underline{\boldsymbol{\mu}_2} \end{bmatrix} \text{ dan } \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\sigma}_{11} & \boldsymbol{\sigma}_{12} \\ \boldsymbol{\sigma}_{21} & \boldsymbol{\sigma}_{22} \end{bmatrix} \tag{2.29}$$

di mana:

$\underline{\boldsymbol{\mu}_k}$ adalah vektor rata-rata kelompok ke-*k*.

$\boldsymbol{\Sigma}$ adalah matriks ragam peragam kelompok 1 dan 2.

Fungsi diskriminan linier ditunjukan dengan persamaan berikut:

$$w = \beta_1 X_1 + \ldots + \beta_p X_p = \underline{\boldsymbol{\beta}}^{\mathrm{T}} \underline{\boldsymbol{X}} \tag{2.30}$$

Fisher menetapkan $\boldsymbol{\beta}^{\mathrm{T}}$ dengan cara memaksimumkan rasio jumlah kuadrat antar kelompok ($\boldsymbol{JK_A}$) dengan jumlah kuadrat dalam kelompok ($\boldsymbol{JK_D}$), di mana $\mathbf{JK_A} = (\mathrm{E}(w/\ \underline{\boldsymbol{x}}_i \in \pi_1) - \mathrm{E}(w/\ \underline{\boldsymbol{x}}_i \in \pi_2))^2 = \left(\underline{\boldsymbol{\beta}}^{\mathbf{T}}\underline{\boldsymbol{\mu}}_{\mathbf{1}} - \underline{\boldsymbol{\beta}}^{\mathbf{T}}\underline{\boldsymbol{\mu}}_{\mathbf{2}}\right)^2$ dan $\mathbf{JK_D} = \mathrm{var}\ (w) = \underline{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{\Sigma}\,\underline{\boldsymbol{\beta}}$ sehingga rasionya menjadi:

$$R = \frac{\boldsymbol{JK_A}}{\boldsymbol{JK_D}} = \frac{\left(\underline{\boldsymbol{\beta}}^{\mathbf{T}}\underline{\boldsymbol{\mu}}_{\mathbf{1}} - \underline{\boldsymbol{\beta}}^{\mathbf{T}}\underline{\boldsymbol{\mu}}_{\mathbf{2}}\right)^2}{\underline{\boldsymbol{\beta}}^{\mathbf{T}}\boldsymbol{\Sigma}\,\underline{\boldsymbol{\beta}}} \tag{2.31}$$

Agar diperoleh titik maksimum, maka $\frac{\partial R}{\partial \underline{\boldsymbol{\beta}}^{\mathrm{T}}} = 0$. Sehingga diperoleh solusi penduga bagi $\underline{\boldsymbol{\beta}}$ yaitu:

$$\underline{\boldsymbol{\beta}} = \boldsymbol{\Sigma}^{-1}(\underline{\boldsymbol{\mu}}_{\mathbf{1}} - \underline{\boldsymbol{\mu}}_{\mathbf{2}}) \tag{2.32}$$

Jika parameter diketahui, maka $\hat{\underline{\boldsymbol{\mu}}}$ dan $\hat{\boldsymbol{\Sigma}}$ diperoleh dari persamaan (2.19) dan (2.20). Namun jika parameter tidak diketahui, maka $\boldsymbol{\beta}^{\mathrm{T}}$ diduga dengan $\mathbf{b}^{\mathrm{T}}$ sebagai berikut:

$$\underline{\boldsymbol{b}} = \boldsymbol{S}^{-1}(\underline{\bar{\boldsymbol{x}}}_{\mathbf{1}} - \underline{\bar{\boldsymbol{x}}}_{\mathbf{2}}) \tag{2.33}$$

dengan

$$\boldsymbol{S} = \sum_{k=1}^{2}\left(\underline{\boldsymbol{x}}_{\boldsymbol{k}}{}^{T}\underline{\boldsymbol{x}}_{\boldsymbol{k}}\right)/(n_1 + n_2 - 2) \tag{2.34}$$

di mana

$\boldsymbol{S}$: matriks ragam peragam contoh

n_1 : ukuran contoh kelompok 1

n_2 : ukuran contoh kelompok 2

$\underline{\bar{\boldsymbol{x}}}_{\mathbf{1}}$: vektor rata-rata contoh kelompok 1

$\underline{\bar{\boldsymbol{x}}}_{\mathbf{2}}$: vektor rata-rata contoh kelompok 2

Aturan klasifikasi berdasarkan fungsi diskriminan Fisher:
$\underline{\boldsymbol{x}}_i$ diklasifikasikan ke dalam kelompok 1 (π_1), jika

$$w \geq m \tag{2.35}$$

atau

$$w - m \geq 0 \tag{2.36}$$

$\underline{\boldsymbol{x}}_i$ diklasifikasikan ke dalam kelompok 2 (π_2), jika

$$w < m \tag{2.37}$$

atau

$$w - m < 0 \tag{2.38}$$

di mana :
w adalah skor diskriminan yang diperoleh dari hasil penjumlahan dari perkalian antara koefisien β dengan nilai peubah prediktor pada setiap obyek.

$$w = \beta_1 X_{1i} + \ldots + \beta_p X_{pi} \tag{2.39}$$

m (*cutting score*) adalah nilai yang digunakan sebagai batas suatu obyek dimasukkan ke dalam kelompok 1 atau 2. Nilai tersebut dibandingkan dengan skor diskriminan. Perhitungan *m* dapat dilakukan dengan menggunakan persamaan sebagai berikut:

$$m = \frac{1}{2}(\bar{w}_1 + \bar{w}_2) \tag{2.40}$$

$$m = \frac{1}{2}\left(\underline{\boldsymbol{\mu}}_\mathbf{1} - \underline{\boldsymbol{\mu}}_\mathbf{2}\right)^T \boldsymbol{\Sigma}^{-1} \left(\underline{\boldsymbol{\mu}}_\mathbf{1} + \underline{\boldsymbol{\mu}}_\mathbf{2}\right) \tag{2.41}$$

Keterangan:

Rata-rata univariat $\bar{w}_1 = \left(\underline{\boldsymbol{\mu}}_\mathbf{1} - \underline{\boldsymbol{\mu}}_\mathbf{2}\right)^T \boldsymbol{\Sigma}^{-1}\, \underline{\boldsymbol{\mu}}_\mathbf{1} = \underline{\boldsymbol{\beta}}^{\boldsymbol{T}} \underline{\boldsymbol{\mu}}_\mathbf{1}$ (2.42)

Rata-rata univariat $\bar{w}_2 = \left(\underline{\boldsymbol{\mu}}_\mathbf{1} - \underline{\boldsymbol{\mu}}_\mathbf{2}\right)^T \boldsymbol{\Sigma}^{-1}\, \underline{\boldsymbol{\mu}}_\mathbf{2} = \underline{\boldsymbol{\beta}}^{\boldsymbol{T}} \underline{\boldsymbol{\mu}}_\mathbf{2}$ (2.43)

(Johnson dan Wichern, 2002).

2.7 Analisis Diskriminan Linier *Robust*

Dalam analisis diskriminan, suatu pengamatan dari beberapa kelompok pengamatan multivariat secara bersama-sama akan membentuk *training set*. *Training set* adalah bagian dari contoh yang siap untuk diteliti di mana obyek sudah diklasifikasifikasikan berdasarkan kelompok menurut informasi awal (informasi prior). Untuk data dalam *training set*, dapat diketahui pengamatan tersebut berasal dari kelompok mana. Aturan diskriminan dibentuk pada dasar *training set* dan digunakan untuk mengklasifikasikan pengamatan baru menjadi satu kelompok. Sebuah metode pengklasifikasian sederhana dan populer yaitu analisis diskriminan Fisher. Fungsi diskriminan Fisher adalah kombinasi linear dari peubah yang diukur. Metode ini bergantung pada rata-rata contoh dan matriks ragam peragam yang dihitung dari kelompok berbeda dan merupakan *training set* (Croux, dkk., 2008).

Analisis diskriminan linier *robust* merupakan perkembangan dari analisis diskriminan linier klasik. Analisis ini digunakan untuk menduga parameter jika terdapat pencilan yang berpengaruh pada data. Pencilan menyebabkan asumsi diskriminan klasik tidak terpenuhi. Jika pendugaan parameter menggunakan metode MLE (*Maximum Likelihood Estimation*) dilakukan dengan analisis diskriminan klasik pada data yang mengandung pencilan, maka dapat mengakibatkan pendugaan parameter yang diperoleh kurang konsisten. Pada masing-masing kelompok perlu diduga vektor rata-rata dan matriks ragam peragam dengan menggunakan rata-rata contoh dan matriks ragam peragam contoh. Namun, rata-rata contoh dan matriks ragam peragam contoh tidak cukup kuat ketika terdapat pencilan pada pengamatan karena pencilan memberikan pengaruh yang terlalu besar pada aturan analisis diskriminan linier klasik (He dan Fung, 2000).

Agar kinerja fungsi diskriminan tetap optimal meskipun terdapat pencilan pada pengamatan, maka digunakan penduga parameter yang tahan dan tidak terpengaruh terhadap pencilan yaitu penduga *robust.* Pada kondisi data tanpa adanya pencilan, penduga *robust* akan sama dengan penduga MLE. Namun, pada kondisi data yang mengadung pencilan, penduga *robust* menghasilkan pendugaan yang lebih konsisten daripada penduga MLE. Tujuan dari penduga *robust* adalah untuk menghasilkan penduga yang efisien pada data yang mengandung pencilan dan bisa meminimumkan bias dalam pendugaan (Rousseuw,dkk., 1998).

Konsep dasar analisis diskriminan linier *robust* yaitu mengganti vektor rata-rata dan matriks ragam peragam dari data awal dengan vektor rata-rata dan matriks ragam peragam yang *robust*. Vektor rata-rata dan matriks ragam peragam yang *robust* kemudian dimasukkan ke dalam fungsi diskriminan klasik sehingga menghasilkan fungsi diskriminan yang *robust*. Oleh karena itu, aturan klasifikasi pada analisis diskriminan klasik tetap menjadi dasar pada analisis diskriminan linier *robust*. Metode *robust* dirancang untuk mengatasi pencilan. Selain itu, metode ini juga dapat digunakan untuk mendapatkan penduga parameter yang *robust*. Salah satu pendekatan untuk mendapatkan penduga parameter dengan prosedur analisis diskriminan linier *robust* adalah menggunakan penduga-S (Croux, dkk., 2008).

2.7.1 Penduga-S

Penduga-S bertujuan untuk menduga parameter fungsi diskriminan linier *robust*. Penduga tersebut dapat dihitung dengan menggunakan empat algoritma yang berbeda yaitu *SURREAL*, *Bisquare*, *Rocke type*, dan *Fast S*. Namun, salah satu perhitungan penduga-S yang lebih mudah yaitu menggunakan algoritma *SURREAL* (*Sufficiently Reliable Regression Algorithm*) yang diperkenalkan oleh Ruppert. Algoritma ini mempermudah perhitungan penduga dengan *breakdown point* yang tinggi dengan pendekatan penyelesaian berdasarkan sub contoh secara acak (Rousseuw,dkk., 1998).

Breakdown point menunjukkan ukuran proporsi pengaruh pencilan pada data di mana suatu metode mampu mengatasi pencilan tersebut dan mempertahankan sifat kekar (*robust*). Nilai tersebut memberikan efek yang besar terhadap pendugaan parameter dengan menghasilkan persentase kecil pada kesalahan pengklasifikasian sehingga ketepatan klasifikasi yang dihasilkan lebih baik. *Breakdown point* yang digunakan pada penduga-S khususnya algoritma *SURREAL* yaitu 0.15, 0.25, dan 0.50. Jika nilai tersebut mendekati 0.50 maka semakin kekar penduga yang dihasilkan terhadap pengaruh pencilan. Namun, hal ini berkebalikan dengan ketepatan klasifikasi yang dihasilkan. Semakin besar *breakdown point*, maka ketepatan klasifikasi yang dihasilkan akan semakin rendah (Croux, dkk., 2008).

2.7.2 Algoritma *SURREAL*

Algoritma *SURREAL* Ruppert dapat digunakan untuk menghitung penduga-S pada data multivariat dengan menggunakan sebaran normal multivariat untuk model awal. Penggunaan sebaran normal multivariat pada model awal bertujuan agar diperoleh kemudahan dan mendapatkan penduga parameter yang memiliki tingkat efisiensi yang tinggi (He dan Fung, 2000).

Menurut Noeryanti (2001), kelebihan algoritma *SURREAL* antara lain:

1. Penduga-S dapat digunakan lebih cepat dan dapat dihitung dalam waktu yang lebih singkat daripada penduga yang lainnya.
2. Penduga-S secara statistik lebih efisien daripada estimasi MVE (*Minimum Volume Ellipsoid*) karena keragamannya kecil akibat keacakan data.

Contoh acak $\underline{\boldsymbol{x}}_1, \underline{\boldsymbol{x}}_2, \ldots, \underline{\boldsymbol{x}}_n$ berukuran p dengan rata-rata $\underline{\boldsymbol{\mu}}$ dan matriks ragam peragam $\boldsymbol{\Sigma}$. Dengan mempertimbangkan jenis dari fungsi ρ, di mana ρ adalah sebuah fungsi obyektif (He dan Fung, 2000):

1. ρ adalah simetri dan kontinu, $\rho(0) = 0$.
2. Jika konstanta $c > 0$ maka ρ tidak mengalami penurunan pada $(0,c)$ dan konstan pada (c,∞), di mana c adalah konstanta.

Penduga-S adalah parameter s yang akan diduga pada pasangan $s(\underline{\hat{\boldsymbol{\mu}}}_S, \hat{\boldsymbol{\Sigma}}_S)$. Untuk data $\{\underline{\boldsymbol{x}}_1, \underline{\boldsymbol{x}}_2, \ldots, \underline{\boldsymbol{x}}_n\}$ yang terdiri dari p peubah pengamatan, penduga-S merupakan penyelesaian dari minimum $s(d_1,\ldots d_n)$. s adalah penduga skala *robust* yang terbentuk dari kombinasi linier jarak *Mahalanobis* yang *robust* yaitu $d_{is} = \left[\left(\underline{\boldsymbol{x}}_i - \underline{\hat{\boldsymbol{\mu}}}_S\right)^T \hat{\boldsymbol{\Sigma}}_S^{-1} \left(\underline{\boldsymbol{x}}_i - \underline{\hat{\boldsymbol{\mu}}}_S\right)\right]^{1/2}$.

Penduga skala *robust* ($\hat{s}$) merupakan penyelesaian dari:

$$\frac{1}{n}\sum_{i=1}^{n} \rho\{(d_i)/\hat{s}\} = b \qquad (2.44)$$

di mana nilai b konstan.

Untuk menentukan vektor $\hat{\boldsymbol{\mu}}_S$ dan matriks $\hat{\boldsymbol{\Sigma}}_S$ yaitu meminimumkan $|\hat{\boldsymbol{\Sigma}}_S|$ dengan batasan (Rousseuw,dkk., 1998):

$$\frac{1}{n}\sum_{i=1}^{n}\rho(d_i) = b \tag{2.45}$$

di mana :
$\hat{\boldsymbol{\mu}}_S$ adalah vektor rata-rata berukuran $p \times 1$ hasil dari penduga-S.
$\hat{\boldsymbol{\Sigma}}_S$ adalah matriks ragam peragam simetri definit positif berukuran $p \times p$ hasil dari penduga-S.
n adalah banyak pengamatan.
p adalah banyak peubah respon.

Breakdown point (ε) untuk penduga-S berkisar antara 0 sampai 0.5 yang dapat ditulis sebagai berikut:

$$\varepsilon = \frac{b}{\rho(c)} \tag{2.46}$$

di mana:
c adalah konstanta untuk mencapai *breakdown point* yang diinginkan dengan mempertimbangkan jenis dari fungsi obyektif (ρ). Nilai c diperoleh dengan menspesifikasi d_i di mana nilai ψ menjadi nol, sehingga $\rho(c)$ diperoleh dari:

$$\rho(c) = \frac{c^2}{6} \tag{2.47}$$

b adalah nilai harapan dari $\rho(d_i)$ yang diasumsikan menyebar multivariat Gaussian yaitu:

$$b = p\chi^2(p+2;c^2)/2 - p(p+2)\chi^2(p+4;c^2)/2c^2+p(p+2)(p+4) \\ \chi^2(p+6;c^2)/6c^4+(c^2/6)\{1-\chi^2(p;c^2)\} \tag{2.48}$$

di mana $\chi^2(p;c^2)$ merupakan sebaran kumulatif untuk χ^2 dengan derajat bebas p.

Dalam menentukan nilai ρ, Tukey menyarankan memakai fungsi obyektif sebagai berikut:

$$\rho(d_i) = \begin{cases} \dfrac{d_i^{\,2}}{2} - \dfrac{d_i^{\,4}}{2c^2} + \dfrac{d_i^{\,6}}{6c^4} & \text{, untuk } |d_i| \leq c \\ \dfrac{c^2}{6} & \text{, untuk } |d_i| > c \end{cases} \tag{2.49}$$

sehingga diperoleh fungsi pengaruh $\psi(d_i)$ yang merupakan turunan pertama dari $\rho(d_i)$ terhadap d_i:

$$\psi(d_i) = \rho'(d_i) = \frac{\partial(\rho(d_i))}{\partial d_i} = \begin{cases} d_i\{1 - \left(\frac{d_i}{c}\right)^2\}^2 & \text{, untuk } |d_i| \leq c \\ 0 & \text{, untuk } |d_i| > c \end{cases} \tag{2.50}$$

Fungsi pembobot $w(d_i) = \frac{\psi(d_i)}{(d_i)}$ maka:

$$w(d_i) = \begin{cases} \{1 - \left(\frac{d_i}{c}\right)^2\}^2 & \text{, untuk } |d_i| \leq c \\ 0 & \text{, untuk } |d_i| > c \end{cases} \tag{2.51}$$

Menurut He dan Fung (2000), langkah pertama dari algoritma *SURREAL* Ruppert yaitu menghitung penduga-S untuk vektor rata-rata $\underline{\boldsymbol{\mu}}_{s1}$ dan $\underline{\boldsymbol{\mu}}_{s2}$ pada semua pengamatan untuk setiap kelompok ($\underline{\boldsymbol{x}}_{\mathbf{ik}}$). Karena ada dua kelompok pengamatan sehingga setiap kelompok dapat dianggap sebagai populasi. Penduga-S diperoleh dengan mengaplikasikan *SURREAL*. Vektor rata-rata penduga-S dapat digunakan untuk memperbaiki rata-rata penduga. Prosedur ini diulang hingga konvergen dengan batasan persamaan (2.37) dan meminimumkan $|\widehat{\boldsymbol{\Sigma}}_S|$, maka diperoleh parameter *robust* penduga-S untuk dua kelompok yaitu (Rousseuw,dkk., 1998):

Vektor rata-rata setiap kelompok:

$$\underline{\hat{\boldsymbol{\mu}}}_{S1} = \frac{\sum_{i=1}^{n1} w(d_{i1})\underline{\boldsymbol{x}}_{i1}}{\sum_{i=1}^{n1} w(d_{i1})} \tag{2.52}$$

$$\underline{\hat{\boldsymbol{\mu}}}_{S2} = \frac{\sum_{i=1}^{n2} w(d_{i2})\underline{\boldsymbol{x}}_{i2}}{\sum_{i=1}^{n2} w(d_{i2})} \tag{2.53}$$

Matriks ragam peragam setiap kelompok:

$$\hat{\boldsymbol{\Sigma}}_{S1} = \frac{\sum_{i=1}^{n1} w(d_{i1})(\underline{\boldsymbol{x}}_{i1}-\underline{\hat{\boldsymbol{\mu}}}_{s1})(\underline{\boldsymbol{x}}_{i1}-\underline{\hat{\boldsymbol{\mu}}}_{s1})^T}{p^{-1}\sum_{i=1}^{n1} d_{i1}\psi(d_{i1})} \tag{2.54}$$

$$\hat{\boldsymbol{\Sigma}}_{S2} = \frac{\sum_{i=1}^{n2} w(d_{i2})(\underline{\boldsymbol{x}}_{i2}-\underline{\hat{\boldsymbol{\mu}}}_{s2})(\underline{\boldsymbol{x}}_{i2}-\underline{\hat{\boldsymbol{\mu}}}_{s2})^T}{p^{-1}\sum_{i=1}^{n2} d_{i2}\psi(d_{i2})} \tag{2.55}$$

Matriks ragam peragam gabungan:

$$\hat{\boldsymbol{\Sigma}}_{S} = \frac{n_1\hat{\boldsymbol{\Sigma}}_{S1}+n_2\hat{\boldsymbol{\Sigma}}_{S2}}{n_1+n_2} \tag{2.56}$$

di mana:
$w(d_{i1})$: pembobot jarak *Mahalanobis* pengamatan ke-*i* kelompok 1
$w(d_{i2})$: pembobot jarak *Mahalanobis* pengamatan ke-*i* kelompok 2
$\underline{\hat{\boldsymbol{\mu}}}_{S1}$: vektor rata-rata untuk kelompok 1
$\underline{\hat{\boldsymbol{\mu}}}_{S2}$: vektor rata-rata untuk kelompok 2
$\hat{\boldsymbol{\Sigma}}_{S1}$: matriks ragam peragam untuk kelompok 1
$\hat{\boldsymbol{\Sigma}}_{S2}$: matriks ragam peragam untuk kelompok 2
n_1 : banyak pengamatan kelompok 1
n_2 : banyak pengamatan kelompok 2
$\underline{\boldsymbol{x}}_{i1}$: pengamatan ke-*i* kelompok ke-1
$\underline{\boldsymbol{x}}_{i2}$: pengamatan ke-*i* kelompok ke-2

Langkah-langkah menduga parameter analisis diskriminan linier *robust* dengan metode penduga-S menggunakan algoritma *SURREAL* adalah (Noeryanti, 2001):

1. Menentukan nilai *breakdown point* (ε) untuk menghitung konstanta *c* dan nilai *b* pada persamaan (2.48).
2. Memilih sub contoh acak berukuran p+1 dari data berukuran *n*.
3. Menghitung penduga awal dari langkah (2) dengan metode MLE *(Maximum Likelihood Estimation*) sehingga diperoleh penduga parameter awal untuk $\underline{\hat{\boldsymbol{\mu}}}$ pada persamaan (2.26) dan $\hat{\boldsymbol{\Sigma}}$ pada persamaan (2.27).
4. Menghitung penduga skala *robust* ($\hat{s}$) pada persamaan (2.44), fungsi pengaruh $\psi(d_i)$ pada persamaan (2.50) dan fungsi pembobot $w(d_i)$ pada persamaan (2.51).
5. Menghitung penduga *robust* dengan memasukkan $\psi(d_i)$ dan $w(d_i)$ ke dalam perhitungan pendugaan, sehingga diperoleh penduga parameter $\underline{\hat{\boldsymbol{\mu}}}_{S1}$ dan $\underline{\hat{\boldsymbol{\mu}}}_{S2}$ pada persamaan (2.52) dan (2.53) serta $\hat{\boldsymbol{\Sigma}}_S$ pada persamaan (2.56).
6. Melakukan pengulangan/iterasi dari langkah (2) sampai (6) hingga diperoleh penduga parameter yang konvergen dengan batasan persamaan (2.45) serta $\left|\hat{\boldsymbol{\Sigma}}_S\right|$ dan $\hat{s}$ yang minimum.

2.7.3 Fungsi Diskriminan Linier *Robust* dan Aturan Pengklasifikasian

Menurut Croux, dkk (2008), dengan memasukkan penduga parameter *robust* hasil dari penduga-S yaitu vektor rata-rata setiap kelompok $\underline{\hat{\boldsymbol{\mu}}}_{S1}$ dan $\underline{\hat{\boldsymbol{\mu}}}_{S2}$ pada persamaan (2.45) dan (2.46) serta matriks ragam peragam gabungan $\hat{\boldsymbol{\Sigma}}_S$ pada persamaan (2.49) ke dalam fungsi diskriminan Fisher pada persamaan (2.26), maka akan diperoleh fungsi diskriminan linier *robust* sebagai berikut:

$$w_r = \beta_{S1}X_1 + \dots + \beta_{Sp}X_\text{p} = \underline{\boldsymbol{\beta}}_S{}^{\text{T}}\underline{\boldsymbol{X}} \qquad (2.57)$$

di mana:

$$\underline{\boldsymbol{\beta}}_S = \hat{\boldsymbol{\Sigma}}_S{}^{-1}\left(\underline{\hat{\boldsymbol{\mu}}}_{S1} - \underline{\hat{\boldsymbol{\mu}}}_{S2}\right) \qquad (2.58)$$

Aturan klasifikasi pada analisis diskriminan linier *robust* berdasarkan aturan klasifikasi pada analisis diskriminan Fisher:
$\underline{\boldsymbol{x}}_i$ diklasifikasikan ke dalam kelompok 1 (π_1), jika

$$w_r \geq m_r \tag{2.59}$$

atau

$$w_r - m_r \geq 0 \tag{2.60}$$

$\underline{\boldsymbol{x}}_i$ diklasifikasikan ke dalam kelompok 2 (π_2), jika

$$w_r < m_r \tag{2.61}$$

atau

$$w_r - m_r < 0 \tag{2.62}$$

di mana :
w_r adalah skor diskriminan linier *robust* yang diperoleh dari hasil penjumlahan dari perkalian antara koefisien β yang *robust* dengan nilai peubah prediktor pada setiap obyek.

$$w_r = \beta_{S1} X_{1\mathrm{i}} + \ldots + \beta_{Sp} X_{\mathrm{pi}} \tag{2.63}$$

m_r adalah *cutting point* yang diperoleh dengan menggunakan penduga parameter *robust* dalam perhitungan sebagai berikut:

$$m_r = \frac{1}{2}\left(\underline{\hat{\boldsymbol{\mu}}}_{S1} - \underline{\hat{\boldsymbol{\mu}}}_{S2}\right)^T \hat{\boldsymbol{\Sigma}}_S^{\ -1}\left(\underline{\hat{\boldsymbol{\mu}}}_{S1} + \underline{\hat{\boldsymbol{\mu}}}_{S2}\right) \tag{2.64}$$

2.8 Ketepatan Fungsi Klasifikasi

Ketepatan fungsi klasifikasi dari analisis diskriminan linier *robust* dapat diketahui dari perhitungan *APER* (*Apparent Error Rate*) yaitu fraksi atau proporsi contoh dari pengamatan pada *training set* yang salah diklasifikasikan oleh fungsi klasifikasi contoh (Johnson dan Wichern, 2002).

Menurut Haykin (1999), langkah awal yang harus dilakukan sebelum menghitung nilai *APER* adalah membagi data menjadi dua bagian. Bagian pertama yaitu *training set* yang digunakan untuk membentuk fungsi klasifikasi dari analisis diskriminan linier *robust,* sedangkan bagian kedua yaitu *testing set* yang digunakan untuk menguji fungsi klasifikasi yang diperoleh dari *training set*. Menurut Hair, dkk (1998), tidak ada aturan baku mengenai pembagian data tersebut. Perhitungan nilai *APER* dilakukan pada *testing set* untuk mengetahui ketepatan klasifikasi yang dihasilkan oleh fungsi klasifikasi yang telah dibentuk dari *training set*.

Untuk mempermudah perhitungan nilai *APER* maka dibuat tabel klasifikasi yang terdiri dari *actual membership* dan *predicted membership* sebagai berikut (Johnson dan Wichern, 2002):

Tabel 2.1. Tabel Klasifikasi

Actual membership	*Predicted membership*			
	π_1	π_2	…	π_g
π_1	n_{11}	n_{12}	…	n_{1g}
π_2	n_{21}	n_{22}	…	n_{2g}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
π_g	n_{g1}	n_{g2}	…	n_{gg}

Keterangan:
n_{11} = jumlah obyek dari π_1 tepat diklasifikasikan sebagai π_1
n_{12} = jumlah obyek dari π_1 salah diklasifikasikan sebagai π_2
n_{ij} = jumlah obyek dari π_i salah diklasifikasikan sebagai π_j ; $i \neq j$
n_{ij} = jumlah obyek dari π_i tepat diklasifikasikan sebagai π_j ; $i = j$
di mana : $i = j = 1, 2, \ldots, g$ dan g adalah jumlah populasi

$$APER = \frac{n_{12} + n_{13} + \ldots + n_{1g} + \ldots + n_{g1} + \ldots + n_{(g-1)g}}{n_1 + n_2 + \ldots + n_g} \quad (2.65)$$

$$= \frac{\text{total obyek yang salah diklasifikasikan}}{\text{total contoh}}$$

2.9 Berat Bayi Lahir Rendah (BBLR)

Berat Bayi Lahir Rendah (BBLR) adalah bayi baru lahir dengan berat badan lahir kurang dari 2500 gram. Neonatus dengan berat badan lahir kurang dari 2500 gram atau sama dengan 2500 gram disebut prematur. Berdasarkan kriteria WHO pada tahun 1961, berat badan bayi saat dilahirkan dibagi dalam dua kelompok yaitu bayi dengan berat badan kurang atau sama dengan 2500 gram

(BBLR) dan bayi dengan berat badan lebih dari 2500 gram (normal). Berdasarkan pengertian di atas, bayi dengan berat badan lahir rendah dapat dibagi menjadi dua golongan yaitu (Bobak, 1999):

1. Prematuritas murni yaitu bayi lahir dengan umur kehamilan kurang dari 37 minggu dan mempunyai berat badan sesuai dengan masa kehamilan atau disebut Neonatus Kurang Bulan-Sesuai Masa Kehamilan (NKB-SMK).
2. Dismaturitas yaitu bayi lahir dengan berat badan kurang dari berat badan seharusnya untuk masa kehamilan, dismatur dapat terjadi pada preaterm, atem, dan post aterm. Dismatur ini dapat juga Neonatus Kurang Bulan-Kecil untuk Masa Kehamilan (NKB-KMK), Neonatus Cukup Bulan-Kecil Masa Kehamilan(NCB-KMK), Neonatus Lebih Bulan-Kecil Masa Kehamilan (NLB-KMK).

Bayi dengan berat lahir rendah pada umumnya kurang mampu menahan tekanan lingkungan yang baru, sehingga dapat berakibat pada terhambatnya pertumbuhan dan perkembangan bahkan dapat mengganggu kelangsungan hidupnya. Selain itu juga akan meningkatkan risiko kesakitan dan kematian bayi karena rentan terhadap infeksi saluran pernafasan bagian bawah, gangguan belajar, masalah perilaku dan lain sebagainya (Depkes RI, 1998).

Secara umum Manuaba (1998) menyatakan bahwa karakteristik BBLR antara lain berat badan kurang dari 2500 gram, panjang badan kurang dari 45 cm, lingkar dada kurang dari 30 cm, lingkar kepala kurang dari 33 cm, umur kehamilan kurang dari 37 minggu, kepala relatif lebih besar, kulit tipis transparan, rambut lanugo banyak, lemak kulit kurang, otot hipotonik lemah, pernafasan tidak teratur, dapat terjadi apnea (gagal nafas), ekstremitas (paha abduksi, sendi lutut atau kaki fleksi lurus), kepala tidak mampu tegak, pernafasan sekitar 45-50 kali per menit, dan frekuensi nadi 100-140 kali per menit.

Faktor-faktor penentu berat badan bayi baru lahir menurut Kardjiati (1987) yaitu:

1. Faktor intrinsik yaitu jenis kelamin, genetika, suku bangsa dan pertumbuhan plasenta.
2. Faktor ibu yang meliputi:
   a. Faktor biologi yaitu umur, paritas, tinggi badan, berat badan prahamil, pertambahan berat badan selama hamil, LILA.

b. Faktor lingkungan yaitu taraf sosial ekonomi, jarak antar kehamilan, penyakit infeksi, kegiatan fisik, perawatan kesehatan, pendidikan, kebiasaan merokok atau minum alkohol dan ketinggian tempat tinggal.

Selain itu, Saraswati (1998) menyatakan bahwa terdapat tiga faktor yang secara langsung dapat mempengaruhi kelahiran bayi dengan berat badan lahir rendah (BBLR) yaitu:

1. Pertambahan berat badan selama hamil
   Pertambahan berat badan normal selama hamil sekitar 10 sampai 12 kg, di mana pada trisemester I penambahan kurang dari 1 kg, trisemester II sekitar 3 kg dan trisemester III sekitar 6 kg.
2. Ukuran Lingkar Lengan Atas (LILA) Wanita Usia Subur (WUS)
   Di Indonesia batas ambang ukuran LILA normal adalah 23.5 cm . Wanita Usia Subur (WUS) dengan ukuran LILA kurang dari 23.5 cm sebelum kehamilan berisiko menderita Kurang Energi Kronis (KEK) yang dapat berakibat melahirkan bayi dengan berat badan lahir rendah (BBLR).
3. Ukuran kadar Hb ibu hamil
   Batas ambang kadar Hb ibu hamil di Indonesia adalah 11 gr/dl. Ibu hamil dengan kadar Hb kurang dari 11 gr/dl beresiko menderita anemia Gizi Besi yang dapat berakibat terjadinya kelahiran bayi dengan berat badan lahir rendah (BBLR). Selain itu juga dapat berisiko pada kematian ibu saat melahirkan.

BAB III
METODOLOGI

3.1 Sumber Data

Data yang digunakan pada penelitian ini adalah data sekunder mengenai berat bayi baru lahir mulai bulan Januari 2012 sampai November 2012 dan peubah-peubah yang mewakili keadaan ibu saat hamil untuk setiap bayi baru lahir yaitu usia ibu, jumlah anak, berat badan ibu prahamil, ukuran Lingkar Lengan Atas (LILA) ibu hamil, kadar Hb ibu hamil dan jarak kelahiran. Data tersebut diperoleh dari buku kohort dari masing-masing bidan desa yang termasuk dalam wilayah Puskesmas Manyar Kabupaten Gresik yaitu desa Peganden, desa Pongangan, desa Sidomukti, desa Manyarejo, desa Sidorukun, desa Leran, dan desa Banjarsari. Alasan penulis meminta data di wilayah Puskesmas Manyar karena berdasarkan data sekunder yang diperoleh dari Dinas Kesehatan Kabupaten Gresik, diantara 32 puskesmas di Kabupaten Gresik yang memiliki jumlah kelahiran bayi dengan berat lahir rendah (BBLR) terbanyak mulai bulan Januari 2012 sampai November 2012 adalah Puskesmas Manyar.

3.2 Peubah Penelitian

Pada penelitian ini, data yang digunakan terdiri dari satu peubah respon Y yang tersaji pada Lampiran 1 dan enam peubah prediktor X_1 sampai X_6 yang tersaji pada Lampiran 2. Peubah respon yang digunakan pada penelitian ini adalah berat bayi baru lahir yang dikelompokkan menjadi dua yaitu kelompok berat bayi lahir normal dengan berat badan lebih dari 2500 gram diberi kode 1 dan kelompok Berat Bayi Lahir Rendah (BBLR) dengan berat badan kurang dari atau sama dengan 2500 gram diberi kode 2. Penentuan kedua kelompok berat bayi baru lahir tersebut berdasarkan kriteria WHO pada tahun 1961. Sedangkan keenam peubah prediktor merupakan peubah-peubah yang mewakili keadaan ibu saat hamil yaitu:

1. Usia ibu (X_1)
   Manuaba (1998) menyebutkan bahwa ibu hamil berusia kurang dari 20 tahun atau lebih dari 35 tahun memiliki risiko tinggi mengalami prematuritas atau melahirkan bayi BBLR.

2. Jumlah anak (X_2)
   Jumlah anak mempengaruhi kondisi kesehatan rahim. Ibu yang telah melahirkan empat kali atau lebih berisiko tinggi mengalami prematuritas karena kondisi rahim yang mulai melemah.
3. Berat badan ibu prahamil (X_3)
   Menurut Kardjiati (1987), berat badan ibu prahamil merupakan salah satu faktor ibu yang dapat mengakibatkan prematuritas.
4. Ukuran LILA ibu hamil (X_4)
   Ibu hamil dengan ukuran LILA kurang dari 23.5 cm berisiko menderita Kurang Energi Kronis (KEK) yang dapat mengakibatkan terjadinya prematuritas.
5. Kadar Hb ibu hamil (X_5)
   Ibu dengan kadar Hb kurang dari 11 gr/dl berisiko menderita anemia gizi yang dapat menyebabkan prematuritas.
6. Jarak kelahiran (X_6)
   Ibu yang melahirkan kurang dari dua tahun setelah kehamilan sebelumnya berisiko mengalami prematuritas karena kondisi rahim yang belum stabil.

Berikut adalah peubah-peubah yang digunakan dalam penelitian ini:

Tabel 3.1 Peubah Penelitian

No.	Peubah	Skala	Keterangan
1.	Y = berat bayi baru lahir	Ordinal	Berskala ordinal dengan ketentuan: 1 = jika termasuk dalam kelompok berat bayi lahir normal 2 = jika termasuk dalam kelompok Berat Bayi Lahir Rendah (BBLR)
2.	X_1 = Usia ibu	Rasio	Satuan: tahun
3.	X_2 = Jumlah anak	Rasio	Satuan: anak
4.	X_3 = Berat badan ibu prahamil	Rasio	Satuan: kg
5.	X_4 = Ukuran LILA ibu hamil	Rasio	Satuan: cm

Lanjutan Tabel 3.1

6.	X_5 = Kadar Hb ibu hamil	Rasio	Satuan: gr/dl
7.	X_6 = Jarak kelahiran	Rasio	Satuan: tahun

3.3 Metode Analisis

Langkah-langkah dalam analisis data yang dilakukan pada penelitian ini adalah sebagai berikut:

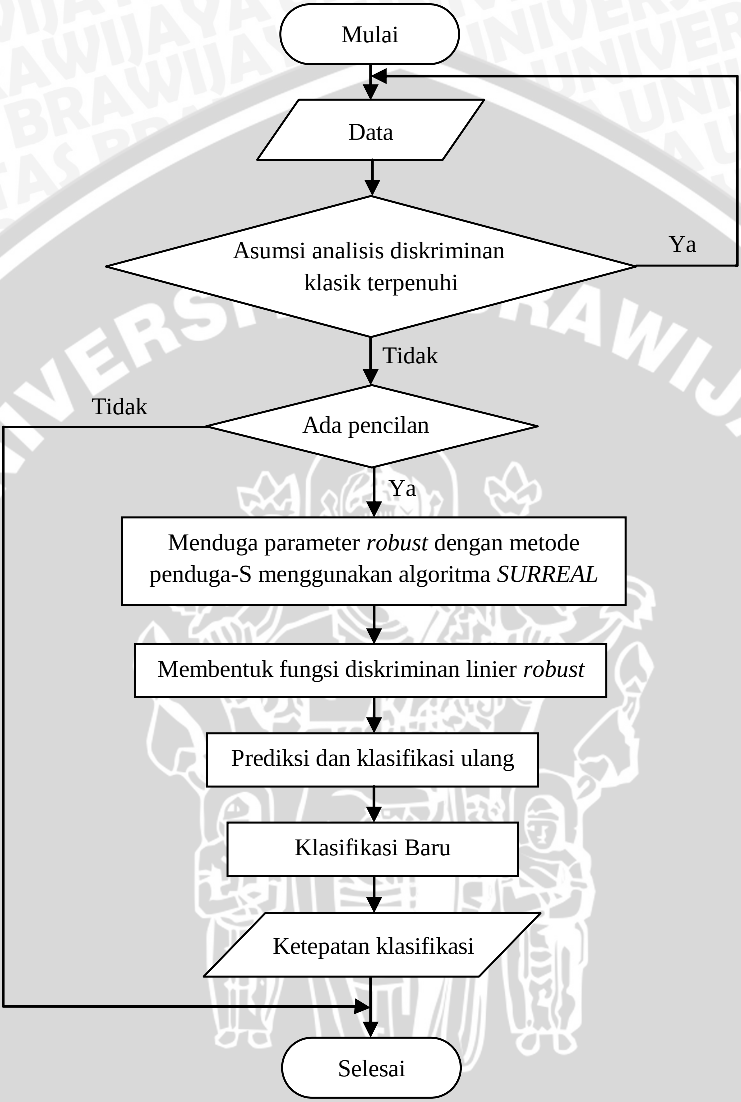
1. Membagi data secara acak menjadi dua bagian yaitu 80% untuk *training set* dan 20% untuk *testing set.*
2. Memasukkan data *training set* dalam bentuk matriks dimana baris merupakan pengamatan dan kolom merupakan peubah.
3. Menguji asumsi analisis diskriminan klasik yaitu asumsi sebaran normal multivariat dengan menggunakan Q-Q plot di mana proses perhitungan dibantu dengan *software* R 2.14.0 dan pengujian asumsi kehomogenan matriks ragam peragam dengan menggunakan statistik uji Box's M pada persamaan (2.14) dengan bantuan *software* SPSS 16. Pengujian asumsi ini dilakukan pada semua peubah prediktor.
4. Menguji perbedaan vektor rata-rata antar kelompok menggunakan statistik uji V-Bartlett pada persamaan (2.17) dengan bantuan *software* SPSS 16.
5. Mengidentifikasi pencilan pada data dengan cara membandingkan jarak *Mahalanobis* pada persamaan (2.8) dengan $\chi^2_{p,0.975}$ yang diperoleh dari tabel χ^2, di mana p adalah banyak peubah prediktor. Identifikasi pencilan dilakukan dengan bantuan *software* R 2.14.0.
6. Menghitung penduga parameter *robust* dengan algoritma *SURREAL* pada metode penduga-S menggunakan bantuan *software* MATLAB R2008a.
   a. Menentukan nilai *breakdown point* (ε) untuk menghitung konstanta c dan nilai b pada persamaan (2.48).
   b. Memilih sub contoh acak berukuran $p+1$ dari data berukuran n.
   c. Menghitung penduga awal dari langkah (2) dengan metode MLE (*Maximum Likelihood Estimation*) sehingga diperoleh

penduga parameter awal untuk $\hat{\underline{\boldsymbol{\mu}}}$ pada persamaan (2.26) dan $\hat{\boldsymbol{\Sigma}}$ pada persamaan (2.27).

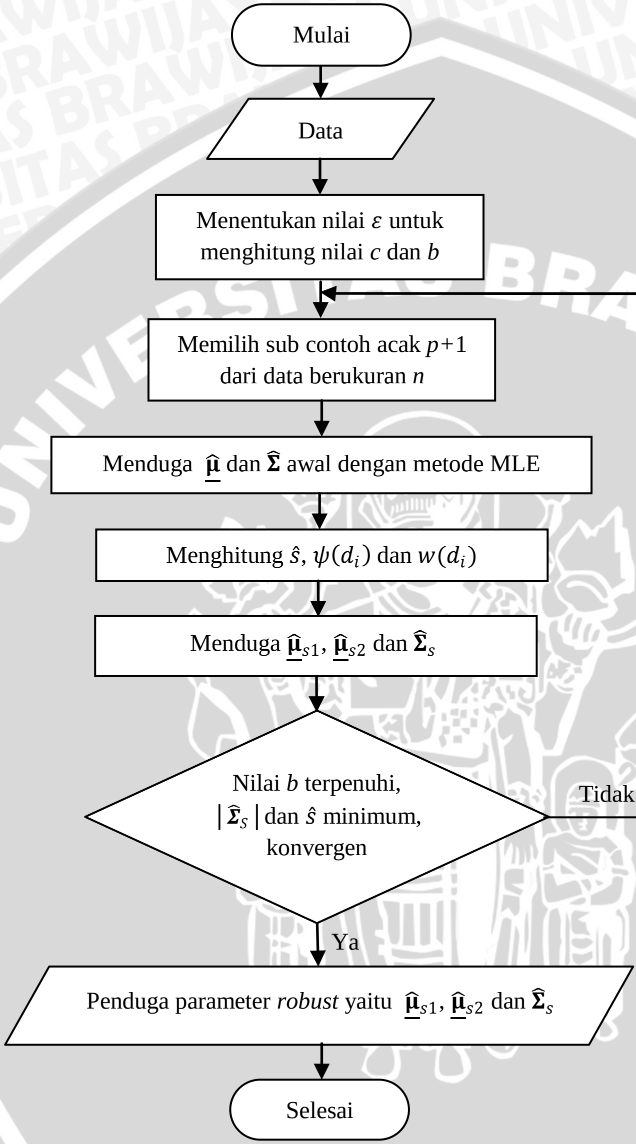
d. Menghitung penduga skala *robust* ($\hat{s}$) pada persamaan (2.44), fungsi pengaruh $\psi(d_i)$ pada persamaan (2.50) dan fungsi pembobot $w(d_i)$ pada persamaan (2.51).
e. Menghitung penduga *robust* dengan memasukkan $\psi(d_i)$ dan $w(d_i)$ ke dalam perhitungan pendugaan, sehingga diperoleh penduga parameter $\hat{\underline{\boldsymbol{\mu}}}_{S1}$ dan $\hat{\underline{\boldsymbol{\mu}}}_{S2}$ pada persamaan (2.52) dan (2.53) serta $\hat{\boldsymbol{\Sigma}}_S$ pada persamaan (2.56).
f. Melakukan pengulangan/iterasi dari langkah (2) sampai (6) hingga diperoleh penduga parameter yang konvergen dengan batasan persamaan (2.45) serta $\left|\hat{\boldsymbol{\Sigma}}_S\right|$ dan $\hat{s}$ yang minimum.

7. Menghitung koefisien $\underline{\boldsymbol{\beta}}_s$ dengan memasukkan penduga parameter *robust* yang diperoleh dari langkah (6) untuk membentuk fungsi diskriminan linier *robust* pada persamaan (2.57). Perhitungan dilakukan dengan bantuan *software* MATLAB R2008a.
8. Melakukan prediksi dan klasifikasi ulang pada *testing set* berdasarkan fungsi diskriminan linier *robust* yang terbentuk dari *training set* dengan bantuan Microsoft Office Excel 2007.
9. Menghitung ketepatan klasifikasi dari analisis diskriminan linier *robust* pada *testing set* menggunakan perhitungan *APER* pada persamaan (2.65) menggunakan bantuan Microsoft Office Excel 2007.

Software statistika yang digunakan untuk perhitungan dan analisis ini adalah R 2.14.0, SPSS 16 dan MATLAB 2008a. Diagram alir penelitian ini disajikan pada Gambar 3.1 dan Gambar 3.2.



Gambar 3.1 Diagram Alir Penelitian



Gambar 3.2 Diagram Alir Algoritma *SURREAL* Metode Penduga-S

BAB IV
HASIL DAN PEMBAHASAN

4.1 Gambaran Umum Berat Bayi Baru Lahir

Menurut WHO pada tahun 1961, berat bayi baru lahir dibagi menjadi dua kelompok yaitu berat bayi lahir normal dengan berat lahir lebih dari 2500 gram dan Berat Bayi Lahir Rendah (BBLR) dengan berat lahir kurang dari atau sama dengan 2500 gram. Jumlah kelahiran BBLR diperkirakan 15% dari seluruh kelahiran di dunia dengan batasan 3.3% sampai 38%. Hal ini lebih sering terjadi di negara-negara berkembang atau sosio-ekonomi rendah. Secara statistik menunjukkan 90% kejadian BBLR di negara berkembang memiliki angka kematian 35 kali lebih tinggi dibandingkan dengan angka kematian bayi dengan berat lahir normal. Angka kejadian BBLR di Indonesia sangat bervariasi antara satu daerah dengan daerah lain yaitu berkisar antara 9% sampai 30%. Target BBLR yang diterapkan pada sasaran program perbaikan gizi menuju Indonesia Sehat tahun 2010 yaitu maksimal 7%.

Pada penelitian ini, analisis diskriminan linier *robust* diterapkan pada kasus berat bayi baru lahir di Puskesmas Manyar Kabupaten Gresik tahun 2012. Analisis ini digunakan untuk klasifikasi ulang berat bayi baru lahir berdasarkan peubah-peubah yang mewakili keadaan ibu saat hamil. Statistik deskriptif digunakan sebagai gambaran awal mengenai berat bayi baru lahir di Puskesmas Manyar Kabupaten Gresik tahun 2012 dan peubah-peubah yang mewakili keadaan ibu saat hamil. Statistika deskriptif untuk berat bayi baru lahir disajikan pada Tabel 4.1 dan statistika deskriptif untuk peubah-peubah yang mewakili keadaan ibu saat hamil disajikan pada Tabel 4.2.

Tabel 4.1 Deskripsi Berat Bayi Baru Lahir di Puskesmas Manyar Kabupaten Gresik Tahun 2012

Kelompok Berat Bayi Baru Lahir	Jumlah	Persentase
Berat Bayi Lahir Normal	276	89,61%
Berat Bayi Lahir Rendah (BBLR)	32	10,39%

Kelompok berat bayi baru lahir merupakan peubah respon berskala non metrik yaitu kategori ordinal. Kelompok tersebut terdiri dari dua kategori yaitu berat bayi lahir normal yang dikategorikan 1 dan Berat Bayi Lahir Rendah (BBLR) yang dikategorikan 2. Berdasarkan Tabel 4.1 dapat disimpulkan bahwa 89.61% bayi yang lahir di Puskesmas Manyar Kabupaten Gresik tahun 2012 termasuk bayi dengan berat lahir normal yaitu bayi dengan berat lahir lebih dari 2500 gram. Sedangkan 10.39% sisanya merupakan persentase bayi dengan berat lahir rendah (BBLR) yaitu bayi dengan berat lahir kurang dari atau sama dengan 2500 gram.

Tabel 4.2 Statistika Deskriptif Peubah-Peubah yang Mewakili Keadaan Ibu saat Hamil

Peubah	Rata-rata	Minimum	Maksimum
Usia ibu	27	17	45
Jumlah anak	1	0	6
Berat badan ibu prahamil	51.7	31	100
Ukuran LILA ibu hamil	25.0	18	34
Kadar Hb ibu hamil	11.3	9.2	22.6
Jarak kelahiran	2.7	0	18

Berdasarkan Tabel 4.2, diketahui bahwa rata-rata usia ibu hamil di Puskesmas Manyar tahun 2012 adalah 27 tahun. Usia ibu hamil yang paling muda adalah 17 tahun dan paling tua 45 tahun. Rata-rata jumlah anak yang dimiliki oleh ibu hamil adalah 1 anak. Jumlah anak paling sedikit yang dimiliki ibu hamil sebelum melahirkan anak yang dikandungnya sekarang adalah 0 anak. Hal ini berarti bahwa ibu hamil tersebut sebelumnya belum memiliki anak dan menanti kelahiran anak pertama yang sedang dikandungnya. Sedangkan jumlah anak paling banyak yang dimiliki ibu hamil sebelum melahirkan anak yang dikandungnya sekarang adalah 6 anak. Rata-rata berat badan ibu prahamil adalah 51.7 kg. Berat badan ibu prahamil paling ringan adalah 31 kg dan paling berat 100 kg. Rata-rata ukuran Lingkar Lengan Atas (LILA) ibu hamil adalah 25.0 cm. Ukuran LILA ibu hamil paling kecil adalah 18 cm dan paling besar 34 cm. Rata-rata kadar Hb ibu hamil adalah 11.3 gr/dl. Kadar Hb ibu hamil paling rendah adalah 9.2 gr/dl dan paling tinggi 22.6 gr/dl. Rata-rata jarak kehamilan anak sebelumnya dengan kelahiran anak yang sekarang adalah 2.7 tahun. Jarak paling dekat antara kehamilan anak sebelumnya dengan kelahiran anak yang sekarang adalah 0 tahun. Hal ini berarti bahwa ibu tersebut sebelumnya belum memiliki anak, dengan kata lain kehamilan dan kelahiran yang sekarang merupakan kehamilan dan kelahiran yang pertama. Jarak paling jauh antara kehamilan anak sebelumnya dengan kelahiran anak yang sekarang adalah 18 tahun.

4.2 Pembagian Data

Proses awal sebelum melakukan klasifikasi obyek dengan analisis diskriminan adalah membagi data menjadi dua bagian yaitu *training set* yang berjumlah 80% dari keseluruhan data dan *testing set* yang berjumlah 20% dari keseluruhan data. *Training set* adalah bagian dari contoh yang digunakan untuk membentuk fungsi klasifikasi dari analisis diskriminan, sedangkan *testing set* adalah bagian dari contoh yang digunakan untuk menguji fungsi klasifikasi yang diperoleh dari *training set*. Pembagian kedua bagian tersebut dilakukan secara acak dengan jumlah yang proporsional untuk masing-masing kelompok. Dalam hal ini, obyek dan kelompok yang dimaksud adalah bayi baru lahir dan kelompok berat bayi baru lahir yang terdiri dari kelompok berat bayi lahir normal dan kelompok Berat Bayi Lahir Rendah (BBLR). Pembagian *training set* dan *testing set* dengan menggunakan aturan 80-20 ditampilkan dalam Tabel 4.3. Sedangkan data *training set* disajikan pada Lampiran 3 dan data *testing set* disajikan pada Lampiran 4.

Tabel 4.3 *Training set* (TS) dan *testing set* (VS)

n	n_1	n_2	n_1TS	n_2TS	n_1VS	n_2VS	nTS	nVS
308	276	32	220	26	56	6	246	62

Keterangan:
n : jumlah keseluruhan data/ ukuran contoh
n_1 : jumlah data pada kelompok 1
n_2 : jumlah data pada kelompok 2
n_1TS : jumlah pengamatan pada *training set* kelompok 1
n_2TS : jumlah pengamatan pada *training set* kelompok 2
n_1VS : jumlah pengamatan pada *testing set* kelompok 1
n_2VS : jumlah pengamatan pada *testing set* kelompok 2

nTS : jumlah seluruh pengamatan pada *training set*
nVS : jumlah seluruh pengamatan pada *testing set*

4.3 Pengujian Asumsi Analisis Diskriminan

Sebelum melakukan klasifikasi ulang obyek yaitu berat bayi baru lahir dengan menggunakan analisis diskriminan, terlebih dahulu harus dilakukan pengujian asumsi dari analisis diskriminan. Pengujian asumsi ini dilakukan pada *training set* karena bagian ini digunakan untuk membentuk fungsi diskriminan. Pada analisis diskriminan, pengujian asumsi dilakukan untuk peubah prediktor karena peubah respon bersifat *fixed* (tidak disyaratkan mengikuti sebaran tertentu). Asumsi yang harus terpenuhi pada analisis diskriminan adalah peubah prediktor menyebar normal multivariat dan kehomogenan matriks ragam peragam antar kelompok.

4.3.1 Uji Asumsi Sebaran Normal Multivariat

Pengujian asumsi sebaran normal multivariat untuk peubah prediktor dapat dilakukan dengan statistik uji d_i^2 (jarak *Mahalanobis*) yaitu membuat plot antara d_i^2 dengan $q_i = \chi^2_{p,\alpha}$, di mana α = (i-0.5)/n dan $\chi^2_{p,\alpha}$ diperoleh dari tabel χ^2. Hipotesis yang digunakan untuk pengujian asumsi tersebut adalah:

H_0 : Data menyebar normal multivariat
H_1 : Data tidak menyebar normal multivariat

Plot kenormalan multivariat (Q-Q plot) disajikan pada Lampiran 5. Dari Q-Q plot dapat dilihat bahwa plot tidak dapat didekati dengan garis lurus (linier), sehingga dapat diputuskan untuk menolak H_0. Selain itu dapat dilihat dari ukuran statistik yaitu ukuran *skewness* ($b_{1,p}$) sebesar 5.665 dan *kurtosis* ($b_{2,p}$) sebesar 0.279. Karena $b_{1,p}\left(\frac{n}{6}\right)$ = 232.265 > χ^2_{56}=74.448 dan $\frac{b_{2,p} - p(p+2)}{\sqrt{(8p(p+2)/n)}}$ = -38.195 < $Z_{0.05}$ = 1.645, maka dapat diputuskan untuk menolak H_0. Hal ini berarti bahwa asumsi sebaran normal multivariat tidak terpenuhi atau dengan kata lain data tidak menyebar normal multivariat.

4.3.2 Uji Asumsi Kehomogenan Matriks Ragam Peragam

Pengujian kehomogenan matriks ragam peragam antar kelompok dapat dilakukan dengan statistik uji Box's M. Karena pada penelitian ini hanya dibatasi untuk dua kelompok, maka hipotesis

untuk pengujian asumsi kehomogenan matriks ragam peragam adalah:
$H_0 : \mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$
$H_1 : \mathbf{\Sigma}_1 \neq \mathbf{\Sigma}_2$

Hasil perhitungan statistik uji Box's M disajikan pada Lampiran 6. Berdasarkan hasil perhitungan, diperoleh nilai Box's M sebesar 50.302. Karena nilai tersebut lebih besar dari $\chi^2_{0.05(21)}$ = 32.67 maka dapat diputuskan untuk menolak H_0. Selain itu, pengambilan keputusan dapat dilihat dari *p-value*. Berdasarkan perhitungan diperoleh *p-value* sebesar 0.0005. Nilai tersebut dibandingkan dengan taraf nyata (α) sebesar 0.05. Karena *p-value* lebih kecil dari α , maka dapat diputuskan untuk menolak H_0. Jadi, dapat disimpulkan bahwa asumsi kehomogenan matriks ragam peragam tidak terpenuhi. Dengan kata lain bahwa matriks ragam peragam antara kelompok berat bayi lahir normal dan kelompok BBLR berbeda atau tidak homogen.

4.4 Perbedaan Vektor Rata-Rata

Pada analisis diskriminan perlu diketahui apakah rata-rata semua peubah prediktor dalam setiap kelompok bernilai sama atau berbeda. Pengujian ini dapat dilakukan dengan statistik uji Wilk's lambda. Karena pada penelitian ini hanya dibatasi untuk dua kelompok, maka hipotesis untuk pengujian perbedaan antar kelompok adalah:
$H_0 : \underline{\boldsymbol{\mu}}_\mathbf{1} = \underline{\boldsymbol{\mu}}_\mathbf{2}$
$H_1 : \underline{\boldsymbol{\mu}}_\mathbf{1} \neq \underline{\boldsymbol{\mu}}_\mathbf{2}$

Hasil perhitungan statistik uji V-Bartlett disajikan pada Lampiran 7. Dari hasil analisis, diperoleh nilai V-Bartlett (χ^2) sebesar 13.902. Nilai tersebut kemudian dibandingkan dengan tabel $\chi^2_{0.05(6)}$ = 12.59. Karena nilai $\chi^2 > \chi^2_{0.05(6)}$, maka dapat diputuskan untuk menolak H_0. Selain itu, pengambilan keputusan juga dapat dilihat dari *p-value* yang dibandingkan dengan taraf nyata (α) sebesar 0.05. Berdasarkan hasil analisis diperoleh *p-value* sebesar 0.031. Karena *p-value* lebih kecil dari α, maka dapat diputuskan untuk menolak H_0. Hal ini berarti bahwa terdapat perbedaan antara vektor rata-rata kelompok berat bayi lahir normal dengan vektor rata-rata kelompok BBLR. Sehingga fungsi diskriminan yang nantinya terbentuk dapat digunakan untuk mengklasifikasikan obyek baru (bayi baru lahir) ke dalam salah satu kelompok tersebut.

4.5 Identifikasi Pencilan

Identifikasi pencilan pada data perlu dilakukan karena pencilan dapat menyebabkan asumsi analisis diskriminan tidak terpenuhi dan klasifikasi obyek yang dihasilkan menjadi kurang tepat. Untuk melakukan identifikasi pencilan pada analisis multivariat dilakukan dengan menggunakan jarak *Mahalanobis* (d_i^2).

Hasil identifikasi pencilan disajikan pada Lampiran 8. Dari hasil analisis dapat dilihat bahwa terdapat beberapa pengamatan yang dapat dikatakan sebagai pencilan di mana nilai $d_i^2 > \chi^2_{0.975(6)} = 14.45$. Berdasarkan hasil analisis tersebut, dapat disimpulkan bahwa terdapat pencilan pada data sebanyak 6.91%. Peneliti memutuskan tidak menghapus pencilan tersebut dan tetap menggunakan dalam analisis selanjutnya karena pencilan tersebut merupakan data yang sebenarnya bukan disebabkan kesalahan penulisan. Selain itu, tidak ada alasan yang kuat untuk menghapus pencilan tersebut.

4.6 Analisis Diskriminan Linier *Robust*

Analisis diskriminan linier *robust* digunakan untuk menduga parameter jika terdapat pencilan yang berpengaruh pada data. Agar kinerja fungsi diskriminan tetap optimal meskipun terdapat pencilan, maka digunakan penduga parameter yang tahan dan tidak terpengaruh terhadap pencilan yaitu penduga *robust*.

4.6.1 Penduga *Robust*

Jika pada data yang mengandung pencilan digunakan metode MLE (*Maximum Likelihood Estimation*) untuk menduga parameter, maka akan diperoleh penduga yang kurang konsisten. Agar diperoleh penduga yang konsisten maka digunakan penduga *robust*. Penduga ini bertujuan untuk menghasilkan penduga yang efisien pada data yang mengandung pencilan dan bisa meminimumkan bias. Salah satu metode yang dapat digunakan untuk mendapatkan penduga *robust* yaitu metode penduga-S. Metode ini dapat menghitung penduga parameter lebih cepat dari penduga yang lainnya. Ada empat algoritma dalam metode tersebut, namun peneliti memilih menggunakan algoritma *SURREAL* karena algoritma ini mempermudah dalam perhitungan penduga dengan pendekatan penyelesaian berdasarkan sub contoh secara acak.

Langkah pertama untuk menduga parameter menggunakan metode penduga-S dengan algoritma *SURREAL* yaitu menentukan besarnya *breakdown point*. Pada algoritma *SURREAL*, *breakdown point* yang digunakan adalah 0.15, 0.25, dan 0.50. Ketiga nilai tersebut akan diterapkan pada data untuk mengetahui nilai c, b, $\hat{s}$ dan ketepatan klasifikasi untuk setiap *breakdown point*. Hasil analisis ditampilkan dalam Tabel 4.4. Untuk mendapatkan penduga yang konsisten maka dilakukan iterasi (N_{samp}) sebanyak 800 kali.

Tabel 4.4 Nilai c, b, $\hat{s}$ dan ketepatan klasifikasi untuk setiap *breakdown point*

Breakdown point	c	b	Skala *robust* ($\hat{s}$)		Ketepatan Klasifikasi
			Kel.1	Kel.2	
0.15	9.7392	2.7559	1.9384	2.0697	80.65%
0.25	7.8953	2.6356	1.9016	2.0686	79.03%
0.50	5.1477	2.2082	1.7852	1.9784	69.35%

Berdasarkan Tabel 4.4, diketahui bahwa semakin besar *breakdown point* maka semakin kecil nilai c, b, dan skala *robust* ($\hat{s}$). Hal tersebut mempengaruhi koefisien dan skor diskriminan linier *robust* serta *cutting point* yang dihasilkan untuk setiap *breakdown point*. Tingkat kekekaran penduga dapat dilihat dari skala *robust* ($\hat{s}$) yang minimum. Berdasarkan hasil analisis pada Tabel 4.4, dapat dibuktikan bahwa semakin besar *breakdown point* maka semakin kekar (*robust*) penduga yang dihasilkan. Perhitungan ketepatan klasifikasi untuk setiap *breakdown point* disajikan pada Lampiran 9. Peneliti memutuskan untuk memilih *breakdown point* sebesar 0.15 agar diperoleh ketepatan klasifikasi yang tinggi. Hal ini sangat penting mengingat tujuan utama peneliti adalah melakukan klasifikasi ulang obyek. Namun karena terdapat pencilan pada data, maka pemilihan *breakdown point* harus tepat agar metode yang digunakan masih dapat mengatasi pencilan dan penduga yang dihasilkan bersifat kekar (*robust*). Dengan memilih *breakdown point* sebesar 0.15 maka metode penduga-S masih dapat mengatasi pencilan sebesar 15%. Pada identifikasi pencilan sebelumnya diketahui bahwa terdapat pencilan pada data sebesar 6.91%, sehingga metode penduga-S masih dapat mengatasi pencilan tersebut dan akan menghasilkan penduga yang kekar (*robust*). Ketika digunakan *breakdown point* yang terlalu besar yaitu 0.25 dan 0.50 untuk pencilan sebesar 6.91%, maka akan dihasilkan ketepatan klasifikasi yang rendah. Hal ini dikarenakan semakin besar *breakdown point* maka semakin besar pula pencilan yang diperbolehkan, sehingga berdampak pada ketepatan klasifikasi yang dihasilkan. Hal-hal tersebut menjadi alasan bagi peneliti memilih *breakdown point* sebesar 0.15 agar diperoleh penduga yang kekar (*robust*) dan ketepatan klasifikasi yang tinggi.

Hasil perhitungan jarak *Mahalanobis* (d_i^2) *robust* , fungsi pengaruh ($\psi(d_i)$) dan fungsi pembobot ($w(d_i)$) untuk setiap kelompok disajikan pada Lampiran 10 sampai Lampiran 15. Berikut adalah hasil penduga *robust* untuk setiap kelompok:

Vektor ata-rata kelompok 1 :

$$\underline{\hat{\boldsymbol{\mu}}}_{S1}=\begin{bmatrix} 26.3114 \\ 0.8047 \\ 51.5504 \\ 25.0487 \\ 11.3170 \\ 2.5258 \end{bmatrix}$$

Matriks ragam peragam kelompok 1:

$$\hat{\boldsymbol{\Sigma}}_{S1}=\begin{bmatrix} 25.2319 & 3.6861 & 10.3459 & 2.8208 & 0.076 & 11.0903 \\ 3.6861 & 1.0606 & 2.4344 & 0.6309 & -0.0039 & 2.3232 \\ 10.3459 & 2.4344 & 88.2912 & 19.6887 & -0.0021 & 5.3209 \\ 2.8208 & 0.6309 & 19.6887 & 6.9146 & -0.1404 & 1.6248 \\ 0.076 & -0.0039 & -0.0021 & -0.1404 & 0.2489 & 0.0215 \\ 11.0903 & 2.3232 & 5.3209 & 1.6248 & 0.0215 & 10.1335 \end{bmatrix}$$

Vektor ata-rata kelompok 2 :

$$\underline{\hat{\boldsymbol{\mu}}}_{S2}=\begin{bmatrix} 28.5494 \\ 1.4148 \\ 49.7353 \\ 24.8239 \\ 11.1682 \\ 2.5514 \end{bmatrix}$$

Matriks ragam peragam kelompok 2:

$$\widehat{\boldsymbol{\Sigma}}_{S2} = \begin{bmatrix} 42.5072 & 6.7775 & 1.274 & -0.7323 & -1.5957 & 10.7098 \\ 6.7775 & 2.8132 & 3.0274 & 0.6409 & -0.2539 & 2.8126 \\ 1.274 & 3.0274 & 66.2531 & 13.6944 & 0.4465 & -4.1972 \\ -0.7323 & 0.6409 & 13.6944 & 6.149 & 0.0465 & -0.5628 \\ -1.5957 & -0.2539 & 0.4465 & 0.0465 & 0.1819 & -0.3804 \\ 10.7098 & 2.8126 & -4.1972 & -0.5628 & -0.3804 & 7.7449 \end{bmatrix}$$

Matriks ragam peragam gabungan:

$$\widehat{\boldsymbol{\Sigma}}_{S} = \begin{bmatrix} 27.0578 & 4.0128 & 9.3871 & 2.4453 & -0.1007 & 11.0501 \\ 4.0128 & 1.2458 & 2.4971 & 0.632 & -0.0304 & 2.3749 \\ 9.3871 & 2.4971 & 85.9619 & 19.0552 & 0.0453 & 4.3149 \\ 2.4453 & 0.632 & 19.0552 & 6.8337 & -0.1207 & 1.3936 \\ -0.1007 & -0.0304 & 0.0453 & -0.1207 & 0.2418 & -0.021 \\ 11.0501 & 2.3749 & 4.3149 & 1.3936 & -0.021 & 9.881 \end{bmatrix}$$

4.6.2 Fungsi dan Prediksi Analisis Diskriminan Linier *Robust*

Konsep dasar dari analisis diskriminan linier *robust* yaitu memasukkan vektor rata-rata dan matriks ragam peragam yang *robust* ke dalam fungsi diskriminan klasik (Fisher), sehingga diperoleh fungsi diskriminan linier *robust*. Oleh karena itu, aturan klasifikasi pada analisis diskriminan klasik tetap menjadi dasar pada analisis diskriminan linier *robust*.

Vektor rata-rata dan matriks ragam peragam yang *robust* serta koefisien diskriminan linier *robust* disajikan pada Lampiran 16. Karena ada dua kelompok pada data, maka hanya satu fungsi diskriminan yang terbentuk. Koefisien untuk masing-masing peubah prediktor dapat diperoleh dari vektor koefisien diskriminan linier *robust* ($\underline{\boldsymbol{\beta}}_S$), yaitu:

$\underline{\beta}_{S1}$ = -0.0878

$\underline{\beta}_{S2}$ = -0.7887

$\underline{\beta}_{S3}$ = 0.0518

$\underline{\beta}_{S4}$ = -0.0541

$\underline{\beta}_{S5}$ = 0.4668

$\underline{\beta}_{S6}$ = 0.2712

sehingga dapat dibentuk fungsi diskriminan linier *robust* sebagai berikut:

$$w_r = \beta_{S1}X_1 + \beta_{S2}X_2 + \beta_{S3}X_3 + \beta_{S4}X_4 + \beta_{S5}X_5 + \beta_{S6}X_6$$

$$w_r = -0.0878X_1 - 0.7887X_2 + 0.0518X_3 - 0.0541X_4 + 0.4668X_5 + 0.2712X_6$$

Fungsi diskriminan linier *robust* tersebut menunjukkan bahwa peubah-peubah prediktor yaitu peubah-peubah yang mewakili keadaan ibu saat hamil menentukan perbedaan antara kelompok berat bayi lahir normal dan kelompok Berat Bayi Lahir Rendah (BBLR). Peubah-peubah yang dimaksud yaitu usia ibu (X_1), jumlah anak (X_2), berat badan ibu pra hamil (X_3), ukuran LILA ibu hamil (X_4), kadar Hb ibu hamil (X_5), dan jarak kelahiran (X_6). Berdasarkan tanda dari koefisien setiap peubah, terlihat bahwa peubah usia ibu, jumlah anak, dan ukuran LILA ibu hamil semuanya bertanda negatif. Namun diantara ketiga peubah tersebut yang memiliki koefisien terbesar adalah peubah jumlah anak. Peubah tersebut merupakan peubah yang berperan paling besar dalam menurunkan skor diskriminan linier *robust* sehingga akan meningkatkan risiko kelahiran bayi BBLR. Dengan kata lain, semakin banyak jumlah anak maka risiko ibu melahirkan bayi BBLR akan semakin besar. Sedangkan untuk peubah usia ibu dan peubah ukuran LILA ibu hamil memiliki koefisien yang kecil, sehingga pengaruh pengurangan dari kedua peubah tersebut terhadap risiko berat dari bayi baru lahir tidak cukup besar. Lain halnya dengan koefisien dari peubah berat badan ibu pra hamil, kadar Hb ibu hamil, dan jarak kelahiran yang semuanya bertanda positif. Namun, diantara ketiga peubah tersebut yang memiliki koefisien paling kecil adalah peubah berat badan ibu pra hamil. Hal ini berarti bahwa peubah tersebut memiliki pengaruh yang tidak cukup besar terhadap risiko berat dari bayi baru lahir. Sehingga peubah kadar Hb ibu hamil dan peubah jarak kelahiran akan meningkatkan skor diskriminan linier *robust* yang mengakibatkan risiko kelahiran bayi dengan berat lahir normal meningkatkan. Dengan kata lain, semakin tinggi kadar Hb ibu hamil dan semakin jauh jarak kelahiran maka risiko melahirkan bayi dengan berat lahir normal akan semakin besar.

Berdasarkan fungsi diskriminan linier *robust* yang terbentuk, dapat dihitung skor diskriminan linier *robust* yang nantinya digunakan untuk memprediksi klasifikasi yang baru. Skor tersebut diperoleh dari hasil penjumlahan dari perkalian antara koefisien *robust* ($\underline{\boldsymbol{\beta}}_S$) dengan nilai-nilai peubah prediktor pada setiap pengamatan. Skor diskriminan linier *robust* untuk setiap pengamatan pada *testing set* disajikan pada Lampiran 17.

Aturan klasifikasi pada analisis diskriminan linier *robust* berdasarkan aturan klasifikasi pada analisis diskriminan Fisher. Untuk mendapatkan prediksi atau klasifikasi yang baru, skor diskriminan linier *robust* dibandingkan dengan nilai *cutting point* sebesar 3.9236. Dengan kata lain, jika bayi baru lahir berdasarkan keadaan ibu saat hamil memiliki skor diskriminan linier *robust* lebih besar atau sama dengan 3.9236 ($w_r \geq 3.9236$), maka bayi tersebut termasuk ke dalam kelompok berat bayi lahir normal. Sebaliknya, bayi baru lahir yang memiliki skor diskriminan linier *robust* kurang dari 3.9236 ($w_r < 3.9236$), maka bayi tersebut termasuk ke dalam kelompok BBLR. Hasil prediksi pengklasifikasian obyek disajikan pada Lampiran 17.

4.7 Ketepatan Fungsi Klasifikasi

Tingkat ketepatan pengklasifikasian sangat menentukan baik atau tidaknya suatu pengklasifikasian obyek. Persentase ketepatan pengklasifikasian obyek dapat dihitung dengan menggunakan *APER*. Perhitungan nilai *APER* dilakukan pada *testing set* untuk mengetahui ketepatan klasifikasi yang dihasilkan oleh fungsi diskriminan linier *robust* yang telah terbentuk dari *training set*. Hasil klasifikasimenurut fungsi diskriminan linier *robust* tidak selalu sama dengan klasifikasi awal. Besarnya kesalahan pengklasifikasian dengan menganggap klasifikasi awal adalah benar merupakan indikator tingkat ketepatan dari fungsi diskriminan linier *robust* yang dihasilkan. Untuk mempermudah perhitungan *APER*, maka dibuat tabel klasifikasi yang terdiri dari *actual membership* dan *predicted membership*. Ketepatan pengklasifikasian obyek yaitu bayi baru lahir berdasarkan fungsi diskriminan linier *robust* ditampilkan dalam Tabel 4.5.

Tabel 4.5 Hasil Pengklasifikasian Bayi Baru Lahir Berdasarkan Fungsi Diskriminan Linier *Robust*

Actual membership	*Predicted membership*		Total
	Kelompok Berat Bayi Normal	Kelompok BBLR	
Kelompok Berat Bayi Normal	48	8	56
Kelompok BBLR	4	2	6
Total	452	10	62

Berdasarkan Tabel 4.4, hasil klasifikasi menunjukkan bahwa dari 56 bayi baru lahir dengan berat lahir normal didapatkan 8 diantaranya diprediksi masuk ke dalam kelompok Berat Bayi Lahir Rendah (BBLR). Sebaliknya, dari 6 bayi baru lahir dengan berat lahir rendah didapatkan 4 diantaranya diprediksi masuk ke dalam kelompok berat bayi lahir normal. Dari hasil tersebut, selanjutnya dapat dihitung ketepatan klasifikasi menggunakan *APER* sebagai berikut:

$$APER = \frac{(4+8)}{62} = 0{,}1935$$

Dari 62 bayi baru lahir yang merupakan bagian dari contoh pada *testing set*, dapat diketahui ketepatan klasifikasi dari analisis diskriminan linier *robust* berdasarkan fungsi diskriminan yang telah terbentuk dari *training set*. Dengan menganggap klasifikasi awal adalah benar, maka diperoleh kesalahan pengklasifikasian berdasarkan hasil dari perhitungan *APER* yaitu sebesar 19.35%. Hal ini berarti bahwa kemampuan analisis diskriminan linier *robust* dalam mengklasifikasikan bayi baru lahir pada data yang mengandung pencilan adalah sebesar 80.65%. Dengan demikian, fungsi diskriminan linier *robust* yang dihasilkan memberikan ketepatan klasifikasi yang cukup tinggi. Sehingga, fungsi tersebut dapat digunakan sebagai fungsi pengklasifikasian bayi baru lahir ke dalam kelompok berat bayi lahir normal atau kelompok Berat Bayi Lahir Rendah (BBLR) berdasarkan peubah-peubah yang mewakili ibu saat hamil. Selanjutnya, berdasarkan fungsi pengklasifikasian ini dapat dipertimbangkan program untuk mengantisipasi atau mencegah terjadinya kelahiran BBLR khususnya bagi ibu hamil yang berisiko melahirkan bayi BBLR.

BAB V
KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil dari penelitian ini, dapat diperoleh beberapa kesimpulan sebagai berikut:

1. Pada identifikasi pencilan diperoleh bahwa terdapat pencilan sebesar 6.91%. *Breakdown point* yang tepat digunakan adalah 0.15 karena metode penduga-S masih dapat mengatasi pencilan dan penduga yang dihasilkan bersifat kekar (*robust*). Selain itu, diperoleh ketepatan klasifikasi yang cukup tinggi.
2. Pada pengklasifikasian ulang bayi baru lahir menggunakan analisis diskriminan linier *robust*, diperoleh fungsi diskriminan linier *robust* yaitu:
   $w_r = -0.0878X_1 – 0.7887X_2 + 0.0518X_3 - 0.0541X_4 + 0.4668X_5 + 0.2712X_6$
   Fungsi tersebut digunakan untuk menghitung skor diskriminan linier *robust* (w_r) dan dibandingkan dengan nilai *cutting point* sebesar 3.9236. Jika $w_r \geq 3.9236$ maka bayi baru lahir termasuk ke dalam kelompok berat bayi lahir normal. Sebaliknya, jika $w_r <$ 3.9236 maka bayi baru lahir termasuk ke dalam kelompok BBLR.
3. Pada pengklasifikasian ulang bayi baru lahir, diperoleh prediksi klasifikasi yang baru dengan kesalahan pengklasifikasian sebesar 19.35%. Hal ini berarti bahwa kemampuan analisis diskriminan linier *robust* dalam mengklasifikasikan bayi baru lahir pada data yang mengandung pencilan adalah sebesar 80.65%.

5.2 Saran

Saran yang direkomendasikan untuk penelitian selanjutnya adalah sebagai berikut:

1. Klasifikasi ulang obyek dapat dilakukan dengan menggunakan analisis diskriminan linier *robust* untuk lebih dari dua kelompok.
2. Pada penelitian selanjutnya dapat digunakan algoritma dari penduga-S selain algoritma *SURREAL* yaitu algoritma *Bisquare*, *Rocke type*, dan *Fast S* untuk mendapatkan penduga parameter yang *robust.*