ANALISIS DISKRIMINAN LINEAR
MENGGUNAKAN *LIKELIHOOD RATIO TEST*

(Skripsi)

Oleh

Meri Handayani

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2016

ABSTRAK

ANALISIS DISKRIMINAN LINEAR MENGGUNAKAN *LIKELIHOOD RATIO TEST*

Oleh

Meri Handayani

Analisis diskriminan adalah suatu teknik peubah ganda yang digunakan untuk mengelompokkan suatu objek ke dalam satu populasi dari beberapa populasi yang ada berdasarkan pengamatan pada beberapa variabel atau karakteristik individu. Penelitian ini bertujuan untuk megkaji analisis diskriminan linear menggunakan *likelihood ratio test*, dan melihat seberapa baik pengklasifikasian dengan menghitung *total probability of misclassification* (TPM) yang mempertimbangkan peluang prior, kemudian diterapkan pada contoh kasus dengan dua populasi. Berdasarkan kajian tersebut diperoleh bahwa pengklasifikasian data semakin baik (kesalahan klasifikasi minimum) ketika peluang prior masing-masing populasi berbeda (dipertimbangkan).

Kata kunci : analisis diskriminan, analisis diskriminan linear, klasifikasi, *likelihood ratio test*, *total probability of misclassification.*

ABSTRACT

LINEAR DISCRIMINANT ANALYSIS USING LIKELIHOOD RATIO TEST

Oleh

Meri Handayani

Discriminant analysis is a technique used multiple variables to classify an object into a population of some existing population based on observations on some variables or characteristics of the individual. This study aims to assess linear discriminant analysis using the *likelihood ratio test*, and to assess how to correct the classification by calculating the *total probability of misclassification* (TPM) that consider opportunities prior, then applied to the case with the two populations. Based on these studies data classification will be better (minimum classification error) when a prior opportunity each distinct population (be considered).

Keywords: discriminant analysis, linear discriminant analysis, classification, *likelihood ratio test*, *total probability of misclassification*.

ANALISIS DISKRIMINAN LINEAR
MENGGUNAKAN *LIKELIHOOD RATIO TEST*

Oleh

Meri Handayani

Skripsi

Sebagai Salah Satu Syarat Untuk Memperoleh Gelar
Sarjana Sains

Pada

Jurusan Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2016

Judul Skripsi	: **ANALISIS DISKRIMINAN LINEAR MENGGUNAKAN *LIKELIHOOD RATIO TEST***
Nama Mahasiswa	: *Meri Handayani*
Nomor Pokok Mahasiswa	: 1117031034
Jurusan	: Matematika
Fakultas	: Matematika dan Ilmu Pengetahuan Alam

MENYETUJUI

1. Komisi Pembimbing

Widiarti, S.Si., M.Si.	**Drs. Rudi Ruswandi, M.Si.**
NIP 19800502 200501 2 003	NIP 19560208 198902 1 001

2. Ketua Jurusan Matematika

Drs. Tiryono Ruby, M.Sc., Ph.D.
NIP 19620704 198803 1 002

MENGESAHKAN

1. Tim Penguji

Ketua	: **Widiarti, S.Si., M.Si.**	....
Sekretaris	: **Drs. Rudi Ruswandi, M.Si.**	....
Penguji Bukan Pembimbing	: **Drs. Eri Setiawan, M.Si.**	....

2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam

Prof. Warsito, S.Si., DEA., Ph.D.
NIP 19710212 199512 1 001

Tanggal Lulus Ujian Skripsi : **4 April 2016**

PERNYATAAN SKRIPSI MAHASISWA

Saya yang bertandatangan di bawah ini :

Nama	: **Meri Handayani**
Nomor Pokok Mahawiswa	: **1117031034**
Jurusan	: **Matematika**

Dengan ini menyatakan bahwa skripsi ini adalah hasil pekerjaan saya sendiri dan semua tulisan yang tertuang dalam skripsi ini telah mengikuti kaidah karya penulisan ilmiah Universitas Lampung.

Bandar Lampung, 4 April 2016

Yang Menyatakan,

Meri Handayani
NPM. 1117031034

RIWAYAT HIDUP

Penulis dilahirkan di Tanjung Karang tanggal 11 Maret 1993, anak pertama dari tiga bersaudara pasangan Bapak Romli dan Ibu Yuli Yanti.

Penulis telah menempuh pendidikan di TK Handayani pada tahun 1999, kemudian menyelesaikan Sekolah Dasar di SD Negeri 4 Gedong Air pada tahun 2005, Sekolah Menengah Pertama Negeri 10 Bandarlampung pada tahun 2008, Sekolah Menengah Atas Negeri 3 Bandarlampung pada tahun 2011.

Penulis terdaftar sebagai mahasiswa S1 Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung pada tahun 2011. Selama menjadi mahasiswa penulis pernah menjadi anggota muda HIMATIKA tahun 2011/2012. Pengurus HIMATIKA sebagai anggota Bidang Eksternal tahun 2012/2013 dan tahun 2013/2014.

Sebagai bentuk pengabdian mahasiswa kepada masyarakat penulis telah mengikuti Karya Wisata Ilmiah (KWI) pada tahun 2012 di Desa Sukabanjar, Tanggamus, Kuliah Praktik (KP) di Badan Pusat Statistika (BPS) Kota Bandarlampung pada tahun 2014, dan Kuliah Kerja Nyata (KKN) yang merupakan mata kuliah wajib untuk strata satu di Desa Sendang Baru Kecamatan Sendang Agung Kabupaten Lampung Tengah, yang dilaksanakan pada tahun 2015.

PERSEMBAHAN

Alhamdulillah hirobbil alamin

Terima kasih sudah menunggu dengan sabar

Teruntuk orang tua tercinta

Ibu Yuli Yanti & Bapak Romli

serta tak lupa teruntuk keluarga, sahabat, teman, dan semua yang mendoakan

SANWACANA

Bismillahirrahmaniirahim

Alhamdulillah, puji syukur kehadirat Allah SWT yang telah melimpahkan berkah dan rahmat-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul "Analisis Diskriminan Linear menggunakan *Likelihood Ratio Test*". Oleh karena itu, penulis ingin mengucapkan terima kasih kepada:

1. Orang Tua tercinta, yang telah memberikan dukungan, doa, dan restu tulus untuk keberhasilan penulis serta telah menunggu dengan sabar.
2. Ibu Widiarti, M.Si. selaku dosen pembimbing utama yang telah meluangkan waktu dari padatnya kesibukkan beliau untuk membimbing, mengoreksi, dan memberi pengarahan kepada penulis hingga skripsi ini selesai.
3. Bapak Rudi Ruswandi, M.Si. selaku dosen pembimbing kedua yang telah banyak membantu, mengoreksi dan memberikan pengarahan dalam proses penyusunan skripsi ini.
4. Bapak Eri Setiawan, M.Si. selaku dosen penguji bukan pembimbing yang memberi penulis masukan dan saran untuk skripsi ini.
5. Bapak Drs. Tiryono Ruby, M.Sc., Ph.D. selaku Ketua Jurusan Matematika FMIPA Universitas Lampung.
6. Bapak Agus Sutrisno, M.Si. selaku pembimbing akademik yang telah memberi nasihat serta pengarahan selama penulis berkuliah.

7. Bapak Prof. Warsito, S.Si., DEA., Ph.D. selaku Dekan FMIPA Universitas Lampung.
8. Dosen, staf dan karyawan Jurusan Matematika FMIPA Universitas Lampung yang telah memberikan ilmu pengetahuan dan bantuan kepada penulis.
9. Kedua adik tersayang, Firman dan Tya yang selalu menghibur dengan segala macam tingkah laku.
10. Tari, Ayu, Putri, Meta, Lala, Dela, Nova, Novia, Ona, Rika, dan Yanti yang selalu memberi keceriaan dalam suka duka penulis.
11. Ica, Dhia (Acong), Anis, Ita, Khairil, Gusti, Sepria, Andzirni, Joko, Wesly, Bang Edo, Guna, Mba Recan, Bunda Lucy, dan pak Drajat yang telah menghibur, memotivasi, dan banyak membantu penulis dimasa-masa penulis berjuang menyelesaikan skripsi ini.
12. Teman - teman seperjuangan Matematika 2011, serta pengurus HIMATIKA. Terima kasih atas keakraban dan kebersamaan selama ini.
13. Muhammad Zulnis Firmansyah yang tidak pernah lelah mendengarkan, menyemangati, serta mendokan penulis dalam keadaan apapun.
14. Untuk orang-orang yang jauh, tapi doanya tidak pernah putus.

Akhir kata, Penulis menyadari bahwa skripsi ini masih jauh dari kesempurnaan, akan tetapi sedikit harapan semoga skripsi yang sederhana ini dapat berguna dan bermanfaat bagi kita semua. Aamiin.

Bandar Lampung,
Penulis,

Meri Handayani

DAFTAR ISI

Halaman

DAFTAR TABEL

Tabel halaman

DAFTAR GAMBAR

I. PENDAHULUAN

1.1 Latar Belakang

Analisis diskriminan merupakan suatu teknik analisis peubah ganda yang digunakan untuk mengelompokkan atau mengklasifikasi suatu objek ke dalam salah satu populasi dari beberapa populasi yang ada berdasarkan pengamatan pada beberapa variabel atau karakteristik individu. Analisis diskriminan adalah salah satu teknik statistik yang digunakan pada hubungan dependensi (hubungan antar variabel yang sudah bisa dibedakan antara peubah respon dan peubah penjelas). Analisis ini digunakan pada kasus dengan peubah respon berupa data kualitatif dan peubah penjelas berupa data kuantitatif. Peubah respon dalam analisis diskriminan berupa data berskala ordinal atau nominal, sedangkan peubah penjelas dalam analisis diskriminan berupa data berskala interval atau rasio. Peubah penjelas ini yang digunakan sebagai pertimbangan dan berpengaruh untuk mengklasifikasikan suatu objek baru ke dalam suatu populasi.

Sebagai metode klasifikasi, fungsi diskriminan dapat digunakan di berbagai bidang terapan, seperti dalam bidang pendidikan. Misalnya pengelola pendidikan tingkat tinggi ingin mengembangkan kriteria penerimaan calon mahasiswa secara objektif. Dari sejumlah variabel bebas, misalnya dalam hal ini adalah nilai raport beberapa mata pelajaran sebagai variabel penjelas, ingin diketahui variabel mana yang dapat dijadikan sebagai variabel peramal keberhasilan studi mahasiswa dan variabel mana yang dapat dijadikan prediksi untuk mengelompokkan mahasiswa ke dalam kelompok berhasil atau gagal (Widiarti, 2003).

Terdapat beberapa metode dalam analisis diskriminan yaitu, analisis diskriminan linear, analisis diskriminan kuadratik, analisis diskriminan fisher, dan analisis diskriminan nonparametrik. Setiap kasus analisis diskriminan memiliki penggunaan yang berbeda dalam menganalisis data. Analisis diskriminan linear digunakan jika data berdistribusi normal multivariat dan setiap kelompoknya memiliki matriks ragam peragam yang homogen. Analisis diskriminan kuadratik digunakan jika data berdistribusi normal multivariat tetapi matriks ragam peragam tidak homogen dalam setiap kelompoknya. Analisis diskriminan fisher digunakan jika data tidak berdistribusi normal multivariat tetapi matriks ragam peragamnya homogen dalam setiap kelompoknya. Analisis diskriminan nonparametrik digunakan jika data tidak berdistribusi normal multivariat dan matriks ragam peragamnya tidak homogen setiap kelompoknya.

Pada analisis diskriminan linear, fungsi diskriminan linear terbentuk dari kombinasi linear variabel-variabel penjelasnya. Metode lain yang dikembangkan dari analisis diskriminan linear yaitu dengan *likelihood ratio test*. *Likelihood ratio test* adalah metode uji perbandingan antara dua model yang bertujuan untuk melihat model mana yang lebih baik untuk diterapkan pada suatu kasus tertentu. Satu model di bawah H_0, dan model lain di bawah H_1. Tes ini didasarkan pada perbandingan nilai maksimum fungsi *likelihood* dari suatu distribusi. Daerah kritis untuk uji hipotesis H_0: $\theta \in \omega$ (kelompok satu) melawan H_1: $\theta \in \omega'$ (kelompok lainnya) dengan ω adalah subset ruang sampel Ω. *Ratio likelihood test* dinotasikan oleh $\lambda = \frac{L(\hat{\omega})}{L(\hat{\Omega})}$ yang kemudian dapat dibandingkan dengan nilai kritis untuk memutuskan apakah akan menolak H_0, di mana dikatakan tolak H_0 ketika nilai rasio $\lambda \leq k$. Namun, kaidah pengklasifikasian berdasarkan indeks atau kriteria apapun tidak selalu bisa diharapkan memiliki ketepatan yang sempurna. Dengan kata lain, dengan penyusunan indeks atau kriteria apapun tetap selalu ada peluang kesalahan klasifikasi. Sedangkan, pengklasifikasian yang baik memiliki peluang kesalahan klasifikasi yang minimum.

Berdasarkan uraian tersebut, penulis tertarik untuk mengkaji lebih dalam tentang analisis diskriminan linear menggunakan *likelihood ratio test* dan melihat seberapa baik pengklasifikasian dengan mencari total peluang kesalahan/ *Total Probability of Misclasification*, kemudian dengan menggunakan *software* R akan dikaji penerapan analisis pada contoh data.

1.2 Perumusan Masalah

Mengingat banyaknya metode pengklasifikasian data yang dapat digunakan, maka fokus penelitian ini adalah mengkaji secara teori mengenai teknik pengklasifikasian suatu data dengan metode analisis diskriminan linear menggunakan *likelihood ratio test*. Analisis diskriminan linear dibatasi dengan data berdistribusi normal ganda dan ragam-peragam homogen, kemudian diaplikasikan pada contoh data dengan dua kelompok.

1.3 Tujuan Penelitian

Berdasarkan latar belakang yang telah dijelaskan sebelumnya, maka tujuan dari penelitian ini adalah:

1. Mengkaji analisis diskriminan menggunakan *likelihood ratio test*.
2. Menghitung *total probability of misclassification* (TPM) dan melihat apakah ada pengaruh nilai peluang prior yang berbeda pada TPM untuk pengklasifikasian dua kelompok.
3. Menerapkan pada contoh data.

1.4 Manfaat Penelitian

Adapun manfaat yang diharapkan dari hasil penelitian ini adalah:

1. Memperdalam pengetahuan mengenai metode pengklasifikasian data, khususnya mengenai metode analisis diskriminan linear.
2. Memberikan motivasi bagi pembaca teori analisis diskriminan agar dapat mengkaji lebih jauh permasalahan yang berhubungan dengan pengklasifikasian data.

II. TINJAUAN PUSTAKA

2.1 Konsep Matriks

Menurut S. Srivastava dan M. Caster (1983), ada beberapa konsep dasar matriks, yaitu sebagai berikut.

2.1.1 Matriks

Misalkan a_{11}, a_{12}, ..., a_{pq} adalah susunan bilangan real dari *pq*. Susunan persegi panjang elemen ini terdiri dari *p* baris dan *q* kolom, ini dinamakan matriks *pxq*.

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1q} \\ a_{21} & a_{22} & \cdots & a_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1p} & a_{2p} & \cdots & a_{pq} \end{bmatrix}$$

Bilangan-bilangan dalam susunan persegi panjang tersebut dinamanan entri dalam matriks.

2.1.2 Transpose Matriks

Jika **A** adalah sebarang matriks *p x q*, maka transpose **A** dinyatakan oleh **A'** dan didefinisikan dengan matriks *q x p* yang kolom pertamanya adalah baris pertama dari **A**, kolom keduanya adalah baris kedua dari **A**, demikian juga dengan kolom selanjutnya merupakan baris selanjutnya dari **A**.

Demikian jika $\mathbf{A}_{p\ x\ q}$, maka $\mathbf{A}'_{q\ x\ p}$.

2.1.3 Invers Matriks

Jika **A** adalah matriks kuadrat, dan jika kita dapat mencari matriks **B** sehingga **AB** = **BA** = **I**, maka **A** dikatakan dapat dibalik (invertible) dan **B** dinamakan invers dari **A**.

2.1.4 Trace Matriks

Suatu matriks yang jumlah baris dan kolomnya sama dikatakan matriks bujur sangkar, jika **A** matriks n x n maka trace (**A**) didefinisikan sebagai berikut :

$$tr(\boldsymbol{A}) = \sum_{i=1}^{n} a_{ii}$$

Dengan a_{ii} adalah unsur diagonal utama.

2.2 Analisis Peubah Ganda dan Vektor Acak

Menurut Johnson dan Wichern (2002), analisis peubah ganda digunakan untuk menganalisa data penelitian yang dikumpulkan dari sejumlah objek dengan setiap objek diukur lebih dari satu peubah respon. Secara umum dalam n buah amatan dilakukan pengukuran p peubah. Data tersebut digambarkan sebagai matriks **X** yang berukuran $n \times p$:

$$\boldsymbol{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \quad (2.1)$$

Matriks **X** memuat data yang terdiri dari seluruh data pengamatan terhadap seluruh peubah penjelasnya.

Pengukuran pada baris ke-i yaitu $x_{i1}, x_{i2}, \dots, x_{ip}$ merupakan pengukuran pada individu yang sama, jika disusun sebagai vektor kolom x_i diperoleh:

$$x_i = \begin{bmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{bmatrix} \quad (2.2)$$

maka x_i disebut sebagai pengamatan vektor acak.

2.3 Distribusi Normal Multivariat

Menurut Johnson dan Wichern (2002), kepekatan normal multivariat merupakan generalisasi dari kepekatan normal univariat untuk dimensi ≥ 2.

Variabel acak X dikatakan berdistribusi normal univariat jika fungsi kepekatan peluangnya adalah :

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]; -\infty < x < \infty \quad (2.3)$$

Misalkan $\left(\frac{x-\mu}{\sigma}\right)^2 = (x-\mu)'(\sigma^2)^{-1}(x-\mu)$ (2.4) adalah fungsi kepekatan normal univariat yang mengukur jarak dari x ke μ dalam satuan standar deviasi. Jarak ini dapat digeneralisasikan untuk vektor pengamatan **x** berukuran p x 1 pada beberapa variabel sebagai:

$$(x-\mu)'\Sigma^{-1}(x-\mu) \qquad (2.5)$$

Vektor $\boldsymbol{\mu}$ berukuran p x l merupakan nilai harapan vektor acak **x** dan matriks Σ berukuran p x p merupakan matriks ragam-peragam. Kita asumsikan matriks simetris Σ adalah definit positif, sehingga persamaan (2.5) merupakan jarak kuadrat dari **x** ke $\boldsymbol{\mu}$.

Kepekatan normal multivariat diperoleh dengan mengganti jarak kuadrat univariat dalam persamaan (2.4) dengan jarak multivariat dalam persamaan (2.5). Sehingga fungsi kepekatan normal *p*-dimensi untuk variabel acak **X** adalah:

$$f(x) = \frac{1}{(2\pi)^{p/2}\,|\Sigma|^{1/2}} exp\left[-\frac{1}{2}\,(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}\,(\boldsymbol{x}-\boldsymbol{\mu})\right] \qquad (2.6)$$

Sehingga dapat ditulis X~N_p ($\boldsymbol{\mu}$, Σ).

2.4 Parameter Distribusi Normal Multivariat

Parameter dari distribusi normal multivariat adalah vektor nilai tengah dan matriks ragam peragam.

2.4.1 Vektor Nilai Tengah

Misalkan $\boldsymbol{x}$ menggambarkan suatu vektor acak dari p peubah pada suatu unit sampel. Jika ada n pengamatan dalam sampel, maka n vektor pengamatan dinotasikan oleh x_1, x_2, ..., x_n. Secara umum dapat dituliskan sebagai:

$$\boldsymbol{x}_i = \begin{bmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{bmatrix} \tag{2.7}$$

Vektor nilai tengah sampel x bisa diperoleh dari rata-rata n vektor pengamatan atau dengan perhitungan rata-rata dari p peubah lainnya secara terpisah (Rencher, 2002).

$$\overline{\boldsymbol{x}} = \frac{1}{n}\sum_{i=1}^{n} x_i = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{bmatrix} \tag{2.8}$$

dengan $\bar{x}_1$ merupakan rata-rata dari n pengamatan pada peubah pertama, $\bar{x}_2$ rata-rata dari peubah kedua, dan seterusnya.

Nilai kemungkinan secara kesuluruhan rata-rata dari $\boldsymbol{x}$ dalam populasi disebut vektor rataan populasi atau nilai harapan dari x. Hal ini didefinisikan sebagai suatu vektor nilai harapan dari setiap peubah.

$$E(\boldsymbol{X}) = E\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix} = \begin{bmatrix} E(x_1) \\ E(x_2) \\ \vdots \\ E(x_p) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} = \boldsymbol{\mu} \tag{2.9}$$

dimana μ_p adalah rata-rata populasi dari p peubah.

Hal ini bisa memperlihatkan bahwa nilai harapan dari $\bar{x}_p$ di $\overline{\boldsymbol{x}}$ adalah μ_p sehingga $E(\bar{x}_p) = \mu_p$. Dengan demikian, nilai harapan $\bar{x}$ adalah:

$$E(\boldsymbol{X}) = E\begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{bmatrix} = \begin{bmatrix} E(\bar{x}_1) \\ E(\bar{x}_2) \\ \vdots \\ E(\bar{x}_p) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} = \boldsymbol{\mu} \tag{2.10}$$

Oleh karena itu, $\overline{\boldsymbol{x}}$ adalah penduga tak bias bagi $\boldsymbol{\mu}$.

2.4.2 Matriks Ragam-peragam

Menurut Raykov dan Marcoulides (2008), matriks ragam peragam merupakan suatu matriks simetris yang berisi ragam pada diagonal utamanya dan koragam pada elemen lainnya. Koefisien ragam menggambarkan sebuah indeks tidak baku dari hubungan linear antara dua peubah penjelas.

Menurut Everitt (2005), ragam populasi dari dua peubah, x_i dan x_j didefinisikan oleh:

$$Cov\left(x_i, x_j\right) = E\left[(x_i - \mu_i)\left(x_j - \mu_j\right)\right] \tag{2.11}$$

Koragam dari x_i dan x_j biasanya dinotasikan oleh σ_{ij}. Jadi, ragam dari peubah x_i sering dinotasikan oleh σ_{ii} dari pada σ_i^2.

Dengan p peubah, x_1, x_2, ..., x_p, ada p ragam dan $\frac{p(p-1)}{2}$ koragam. Secara umum, perhitungan ini dihasilkan dari suatu $p \times p$ matriks simetris $\sum$, yaitu:

$$\begin{aligned}
\sum &= E[(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})'] \\
&= E\left(\begin{bmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \\ \vdots \\ X_p - \mu_p \end{bmatrix} [X_1 - \mu_1 \quad X_2 - \mu_2 \quad \cdots \quad X_p - \mu_p]\right) \\
&= \begin{bmatrix} E(X_1 - \mu_1)^2 & E(X_1 - \mu_1)(X_2 - \mu_2) & \cdots & E(X_1 - \mu_1)\left(X_p - \mu_p\right) \\ E(X_2 - \mu_2)(X_1 - \mu_1) & E(X_2 - \mu_2)^2 & \cdots & E(X_2 - \mu_2)\left(X_p - \mu_p\right) \\ \vdots & \vdots & \ddots & \vdots \\ E\left(X_p - \mu_p\right)(X_1 - \mu_1) & E\left(X_p - \mu_p\right)(X_2 - \mu_2) & \cdots & E\left(X_p - \mu_p\right)^2 \end{bmatrix} \\
&= \begin{bmatrix} var(x_1) & cov(x_1, x_2) & \cdots & cov\left(x_1, x_p\right) \\ cov(x_2, x_1) & var(x_2) & \cdots & cov\left(x_2, x_p\right) \\ \vdots & \vdots & \ddots & \vdots \\ cov\left(x_p, x_1\right) & cov\left(x_p, x_2\right) & \cdots & var\left(x_p\right) \end{bmatrix}
\end{aligned}$$

$$= \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \dots & \sigma_{pp} \end{bmatrix} \tag{2.12}$$

dengan $\sigma_{ij} = \sigma_{ji}$. Matriks ini biasanya disebut matriks ragam peragam atau matriks koragam. Matriks $\sum$ diduga oleh matriks S.

S adalah penduga matriks ragam peragam kelompok ke-*i* yang didefinisikan oleh:

$$S = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})' \tag{2.13}$$

dengan $x_i' = [x_{i1}, x_{i2}, \dots, x_{ip}]$ adalah vektor pengamatan untuk *i* pengamatan.

Diagonal utama dari matriks **S** berisi ragam dari peubah lainnya.

2.5 Kombinasi Linear

Pada analisis diskriminan, fungsi diskriminan terbentuk dari kombinasi linear variabel-variabel penjelasnya.

Menurut Howard dan Romes (2003) sebuah vektor Y dinamakan kombinasi linear dari vektor-vektor x_1, x_2, ..., x_n jika vektor tersebut dapat diungkapkan dalam bentuk

$$Y = b_1 x_1 + b_2 x_2 + \dots + b_n x_n = \boldsymbol{b'x}$$

Dengan b_1, b_2, ..., b_n adalah skalar.

2.6 Jarak Mahalanobis

Menurut Seber (1983), jarak mahalanobis adalah ukuran jarak yang didasarkan pada korelasi antar variabel-variabel, khususnya invers matriks kovariansi. Kuadrat jarak mahalanobis antara dua vektor x_i dan x_j, dengan matriks kovariansi $\sum$, adalah

$$d_{ij}^2 = \sqrt{\left(x_i - x_j\right)' \sum^{-1} \left(x_i - x_j\right)}$$

2.7 Analisis Diskriminan

Menurut Johnson & Wichern (2002) analisis diskriminan merupakan suatu teknik peubah ganda yang digunakan untuk memisahkan pengamatan atau objek ke dalam kelompok atau himpunan yang berbeda dan untuk mengklasifikasikan objek baru ke dalam salah satu kelompok yeng telah ditentukan sebelumnya.

Analisis diskriminan adalah salah satu teknik statistik yang bisa digunakan pada hubungan dependensi (hubungan antarvariabel dimana sudah bisa dibedakan mana variabel respon dan mana variabel penjelas). Lebih spesifik lagi, analisis diskriminan digunakan pada kasus dimana variabel respon berupa data kualitatif dan variabel penjelas berupa data kuantitatif. Ide dasar dari analisis diskriminan adalah untuk menghasilkan aturan yang memungkinkan kita untuk memperkirakan dari populasi mana pengamatan tersebut lebih mungkin berasal.

Menurut Giri (2004), ide dasar analisis diskriminan yaitu dari pengelompokan suatu individu ke salah satu dari beberapa populasi berbeda yang ada berdasarkan pengamatan pada beberapa karakter individu. Misalkan diberikan k populasi berbeda π_1, ..., π_k, akan diklasifikasikan suatu individu dengan pengamatan $x = (x_1, \dots, x_p)'$ ke salah satu dari populasi π_1, ..., π_k.

Analisis diskriminan merupakan suatu fungsi yang terdiri dari kombinasi linear dari dua atau lebih peubah bebas yang paling baik dalam membedakan antara dua kelompok atau lebih (Sartono, 2003). Jika **X** merupakan peubah acak berdimensi p-variat dan b_k merupakan koefesien diskriminan yang akan diduga, maka fungsi diskriminan dapat dituliskan:

$$Y_k = b_{k1}X_1 + b_{k2}X_2 + \cdots + b_{kp}X_p = \boldsymbol{b}_k'\boldsymbol{X} \qquad (2.14)$$

Dengan

Y_k	= nilai diskriminan ke-k dengan k = 1,2, ..., s dan s ≤ min (n-1,p)
p	= jumlah peubah penjelas
n	= jumlah populasi
b	= koefisien diskriminan
X	= peubah penjelas

2.8 Analisis Diskriminan Linear

Analisis diskriminan linear merupakan metode analisis diskriminan yang digunakan pada kondisi data berdistribusi normal multivariat dan asumsi keidentikan/homogen matriks ragam peragam antar kelompok terpenuhi. Fungsi diskriminan linear merupakan kombinasi linear variabel-variabel asal yang akan menghasilkan cara terbaik dalam pemisahan kelompok. Banyaknya fungsi diskriminan yang terbentuk secara umum tergantung dari *g* kelompok dan *p* banyaknya variabel bebas.

Misalkan dua populasi normal peubah ganda mempunyai matriks ragam peragam sama ($\sum_1 = \sum_2$), serta μ_1 dan μ_2 masing-masing merupakan nilai harapan dari populasi π_1 dan π_2, di mana π_1 = populasi 1 dan π_2 = populasi 2, maka:

pilih π_1 jika **b'x** – h > 0 dan (2.15)

pilih π_2 selainnya

dimana

$\mathbf{b} = \sum^{-1} (\mu_1 - \mu_2)$

$h = \frac{1}{2}(\mu_1 - \mu_2)\sum^{-1}(\mu_1 + \mu_2)$

fungsi **b'x** disebut koefisien fungsi diskriminan linear pada x (Johnson & Wichern, 2002).

2.9 Asumsi Analisis Diskriminan Linear

Beberapa asumsi yang mendasari fungsi diskriminan linear adalah:

2.9.1 Uji Distribusi Normal Multivariat

Asumsi kenormalan peubah ganda dibutuhkan untuk uji signifikan pembeda peubah dan fungsi diskriminan. Pengujian data berdistribusi normal multivariat dapat dilakukan dengan menggunakan plot jarak mahalanobis (D_i^2) dan khi-kuadrat $\left(\mathcal{X}_p^2\left((i - 0{,}5) \,/\, n\right)\right)$ (Johnson dan Wichern, 2002).

Setiap vektor pengamatan dapat dihitung jarak mahalanobisnya (D_i^2) dengan persamaan:

$$d_i^2 = (\boldsymbol{x}_i - \overline{\boldsymbol{x}})'\boldsymbol{S}^{-1}(\boldsymbol{x}_i - \overline{\boldsymbol{x}}) \qquad i = 1,2, ..., n \qquad (2.16)$$

di mana

$\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ adalah sampel pengamatan

$\mathbf{S}^{-1}$ adalah kebalikan (*inverse*) matrik kovarians **S**.

Kemudian $d_i^2 = d_1^2, d_2^2, \ldots, d_n^2$ dibuat plot d_i^2 dengan nilai Khi-Kuadrat $\left(\chi_p^2(i - {}^1\!/_2)\right)/n$, di mana i = urutan = 1, 2, ...n, dan p = banyaknya peubah diurutkan dari kecil ke besar. Bila hasil *plot* dapat didekati dengan garis lurus atau berada di sekitar garis lurus, maka dapat disimpulkan bahwa data menyebar secara normal ganda. Jika asumsi kenormalan tidak dipenuhi, maka kita dapat melakukan pemilihan jenis transformasi terhadap data tersebut.

Selain itu, statistik uji Shapiro Wilk juga dapat digunakan untuk menguji kenormalan dengan hipotesis berdasarkan sampel acak berukuran n, $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$. Secara umum, untuk pengujian data berdistribusi normal mutivariat, digunakan hipotesis:

$H_0 = X_1, X_2, \ldots, X_n$ berdistribusi multivariat normal

$H_1 = X_1, X_2, \ldots, X_n$ tidak berdistribusi multivariat normal

Pengujian asumsi yang digunakan adalah *Shapiro-Wilk's Test.* Uji Statistik Shapiro-Wilk didasarkan pada suatu sampel acak berukuran n, x_1, ..., x_n yang didefiniskan sebagai:

$$W_X = \frac{\tilde{\sigma}_X^2}{S_X^2} \qquad (2.17)$$

dengan $S_X^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2$ dan $\tilde{\sigma}_X^2 = [\sum_{i=1}^{n} a_i\,(x_{n-i+1} - x_i]^2$ dimana a_i adalah anggota ke-i dari koefisiean uji shapiro wilk.

Uji ini akan tolak H_0 dengan suatu ukuran taraf nyata α jika $W_X < k_\alpha$ dengan k_α merupakan persentil $100\alpha\%$ dari distribusi W_X (Alva dan Estrada, 2009).

2.9.2 Uji Kehomogenan Matriks Ragam-peragam

Selain uji kenormalan peubah ganda, uji kehomogenan suatu matriks ragam-peragam juga dibutuhkan untuk uji signifikan pembeda peubah dan fungsi diskriminan. Untuk menguji kehomogenan matriks ragam ($\sum$) antar kelompok, dapat digunakan hipotesis:

$H_0 : \sum_1 = \sum_2 = ... = \sum_k$

$H_1 : \sum_i \neq \sum_j$ (sedikitnya ada dua kelompok yang berbeda) $i \neq j = 1,2, ...,k$

Statistik uji yang digunakan adalah statistik Box's M, yaitu:

$$-2\ ln\ \lambda^* = (n-k)\ ln\ |\ W/(n-k)\ |\ - \sum (ni - 1)\ ln\ |Sj| \qquad (2.18)$$

dengan:

$$\lambda^* = \frac{\prod |S_j|^{(n_j-1)/2}}{|W/(n-k)|^{(n-k)/2}}$$

k = banyaknya kelompok

$W/(n-k)$ = matriks ragam-peragam dalam kelompok gabungan

S_j = matriks ragam-peragam kelompok ke-j

Bila hipotesis nol benar, maka:

$(-2\ ln\ \lambda^*)/b$ akan mengikuti sebaran F dengan derajat bebas v_1 dan v_2 pada taraf nyata α, dimana:

$$v_1 = (1/2)(k-1)p(p+1)$$

$$v_2 = (v_1+2)/(a_2 - a_1^2)$$

$$b = v_1/(1-a_1-v_1/v_2)$$

dengan,

$$a_1 = \frac{2p^2+3p-1}{6(k-1)(p+1)}\left[\sum\frac{1}{(n_j-1)} - \frac{1}{(n-k)}\right]$$

$$a_2 = \frac{(p-1)(p-2)}{6(k-1)}\left[\sum\frac{1}{(n_j-1)^2} - \frac{1}{(n-k)^2}\right]$$

p = jumlah variabel penjelas dalam fungsi diskriminan

Karena itu, apabila $(-2\ ln\ \lambda^*)/b > F_{v_1,v_2,\alpha}$ maka H_0 ditolak dan dapat disimpulkan bahwa terdapat kelompok yang memiliki matriks ragam-peragam yang tidak homogen (Mattjik & Sumertajaya, 2011).

2.10 Metode Kemungkinan Maksimum Likelihood

Menurut Rencher (2002), ketika suatu distribusi seperti normal multivariat diasumsikan untuk semua populasi, nilai dugaan bagi parameter sering diperoleh dengan metode kemungkinan maksimum likelihood (*maximum likelihood estimation*). Vektor pengamatan $\boldsymbol{x_1}$, $\boldsymbol{x_2}$, ..., $\boldsymbol{x_n}$ dianggap diketahui dan nilai $\boldsymbol{\mu}$ dan $\sum$ dicari dengan memaksimumkan densitas bersamanya yang disebut fungsi likelihood, yaitu:

$$L\ (\mathbf{X};\ \boldsymbol{\mu}, \textstyle\sum) = \prod_{i=1}^{n} f\ (\boldsymbol{x};\ \boldsymbol{\mu}, \textstyle\sum)$$

$$= \frac{1}{(2\pi)^{np/2}\,|\Sigma|^{n/2}} \exp\left[-\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{x_i} - \boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x_i} - \boldsymbol{\mu})\right]$$

$$= |\,2\pi\,|^{-np/2}\,|\Sigma|^{-n/2} \exp\left[-\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{x_i} - \boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x_i} - \boldsymbol{\mu})\right] \tag{2.19}$$

Untuk normal multivariat, penduganya adalah:

$$\hat{\boldsymbol{\mu}} = \overline{\boldsymbol{x}}$$

$$\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x_i} - \overline{\boldsymbol{x}})(\boldsymbol{x_i} - \overline{\boldsymbol{x}})'$$

$$= \frac{1}{n}\boldsymbol{W}$$

$$= \frac{n-1}{n}\boldsymbol{S} \tag{2.20}$$

dengan $\boldsymbol{W} = \sum_{i=1}^{n}(\boldsymbol{x_i} - \overline{\boldsymbol{x}})(\boldsymbol{x_i} - \overline{\boldsymbol{x}})'$ dan **S** adalah matriks varian kovarian sampel yang didefiniskan:

$$\mathbf{S} = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{21} & s_{22} & \cdots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \cdots & s_{pp} \end{bmatrix} \tag{2.21}$$

2.11 *Likelihood Ratio Test*

Misalkan $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ melambangkan n peubah acak independen yang memiliki masing-masing fungsi kepekatan peluang $f_1(x_i\,;\ \theta_1, \theta_2, \ldots, \theta_m)$, dengan i = 1, 2, ..., n. Himpunan yang terdiri dari semua titik parameter $(\theta_1, \theta_2, \ldots, \theta_m)$ dinotasikan oleh Ω, yang disebut dengan ruang sampel dari semua observasi $\boldsymbol{x}$ yang mungkin. Misalkan ω menjadi sebuah subset dari ruang sampel Ω.

Misalkan hipotesis H_0: $(\theta_1, \theta_2, \ldots, \theta_m) \in \omega$

dan H_1: $(\theta_1, \theta_2, \ldots, \theta_m) \in \omega'$.

Definisi fungsi likelihood maksimum:

$$L(\omega) = \prod_{i=1}^{n} f_i(x_i\,;\ \theta_1, \theta_2, \dots, \theta_m), \quad (\theta_1, \theta_2, \dots, \theta_m) \in \omega$$

dan

$$L(\Omega) = \prod_{i=1}^{n} f_i(x_i\,;\ \theta_1, \theta_2, \dots, \theta_m), \quad (\theta_1, \theta_2, \dots, \theta_m) \in \Omega$$

Misalkan $L(\widehat{\omega})$ dan $L(\widehat{\Omega})$ maksimum, yang diasumsikan ada dari dua fungsi kemungkinan. Rasio dari $L(\widehat{\omega})$ dan $L(\widehat{\Omega})$ disebut rasio kemungkinan (*likelihood ratio*) dan dinotasikan oleh

$$\lambda\,(x_1, x_2, \dots, x_n) = \lambda = \frac{L(\widehat{\omega})}{L(\widehat{\Omega})} \tag{2.22}$$

Menurut Hogg dan Craig (1978), tes dikatakan menolak hipotesis H_0 ketika nilai rasio λ ini kecil, katakan $\lambda \leq k$.

2.12 *Total Probability of Misclassification* (TPM)

Menurut Giri (2004), misalkan seluruh ruang berdimensi p dari X dilambangkan E^p, akan ditentukan aturan untuk membagi E^p ke dalam k daerah yaitu L_1, ..., L_k, sehingga jika **x** jatuh di L_i di mana L_i = 1, ..., k, maka x akan diklasifikasikan sebagai anggota populasi π_i. Namun ada kemungkinan bahwa x yang seharusnya merupakan anggota populasi π_i diklasifikasikan menjadi anggota populasi π_j.

Peluang kesalahan klasifikasi suatu individu dengan pengamatan x masuk ke populasi π_j tetapi seharusnya masuk ke populasi π_i adalah

$$p(j|i, L) = \int_{L_j} f_i(x)\, dx \tag{2.23}$$

Misalkan p_i dilambangkan sebagai peluang prior dari π_i, $i = 1, \dots, k$. Jika p_i diketahui, dapat ditetapkan rata-rata kesalahan klasifikasi suatu individu. Karena peluang yang menggambarkan suatu pengamatan dari populasi π_i adalah p_i dan pengelompokan ke dalam populasi π_i secara tepat (tidak terjadi kesalahan klasifikasi) dengan bantuan dari aturan daerah L, peluang klasifikasi dapat dituliskan dengan $p_i \; p(i|i, L), i = 1, \dots, k$. Sedangkan peluang yang menggambarkan suatu pengamatan x masuk ke populasi π_j tetapi seharusnya masuk ke populasi π_i $(i \neq j)$, dengan cara yang sama peluang kesalahan klasifikasi dapat dituliskan dengan $p_i \, p(j|i, L),, i = 1, \dots, k$.

Sehingga *Total Probality of Misclassification* (TPM) yang merupakan total dari peluang kesalahan pengklasifikasian L dengan mempertimbangkan peluang prior $p = (p_i, \dots, p_k)$ dituliskan sebagai berikut:

$$\sum_{i=1}^{k} p_i \sum_{j=1, j\neq i}^{k} p(j|i, L) \qquad (2.24)$$

`

III. METODOLOGI PENELITIAN

3.1 Waktu dan Tempat Penelitian

Penelitian ini dilakukan pada semester genap tahun akademik 2015/2016, bertempat di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.

3.2 Data Penelitian

Dalam penelitian ini, data yang digunakan diambil dari Jurnal *Analisis Diskriminan dalam Penelitian Ekonomi* oleh Purwo Susongko dan Inayah Adi Sari (2012) tentang rumah tangga yang mengunjungi suatu tempat rekreasi terkenal dengan data yang sedikit diubah. Rumah tangga dibagi dua, yaitu kelompok 1 yang dua tahun terakhir mengujungi tempat rekreasi dan kelompok 2 yang tidak. Data sebanyak 42 rumah tangga ini, diukur oleh tiga peubah penjelas, yaitu X_1 = penghasilan/pendapatan tahunan keluarga ($ 000), X_2= banyaknya anggota rumah tangga (beberapa orang), X_3 = usia kepala rumah tangga (tahun). Pada jurnal tersebut dibahas mengenai pengklasifikasian dengan analisis diskriminan menggunakan *cross validation*, sedangkan pada penelitian ini, penulis hanya menggunakan data dari skripsi tersebut untuk menerapkan suatu pengklasifikasian dengan analisis diskriminan menggunakan *likelihood ratio test*.

Pengamatan dilakukan sebanyak tiga kali. Pertama, ukuran n_1 = n_2, dengan n_1 = n_2 = 21. Yang kedua ukuran $n_1 < n_2$, dengan n_1 sebanyak 15 dan n_2 sebanyak 21. Dan yang ketiga n_1 > n_2, dengan n_1 sebanyak 21 dan n_2 sebanyak 18. Untuk ukuran data yang kedua dan ketiga diperoleh dengan melakukan sampling pada kelompok 1 dan kelompok 2 pengamatan pertama (tentang rumah tangga yang mengunjungi suatu tempat rekreasi terkenal) dengan bantuan *software* R.

Pada penelitian ini, alasan dilakukan sampling adalah untuk mendapatkan nilai peluang prior yang berbeda ($p_1 \neq p_2$), nilai prior ini diperoleh dari $\frac{n_i}{N}$. Nilai p_1 dan p_2 dibutuhkan untuk melihat apakah ada pengaruh besarnya *total probability of misclassification* pada saat peluang prior kedua kelompok berbeda. *Total probability of misclassification* pada penelitian ini digunakan untuk melihat seberapa baik pengklasifikasian data yang telah dilakukan dengan metode analisis diskriminan menggunakan *likelihood ratio test*. Secara lengkap data tersaji pada Lampiran 2.

Struktur data pada analisis diskriminan linear ini terdiri dari dua populasi yang disimbolkan dalam Y_1 dan Y_2. Nilai pengamatan pada kelompok ke-*i* untuk pengulangan ke-*j* dan variabel ke-*k* disimbolkan dalam X_{ijk}, dan Pengamatan ke-*j* pada kelompok ke-*i* disimbolkan dalam j_i.

Struktur data pada analisis diskriminan tersaji pada Tabel 3.1.

Tabel 3.1 Struktur Data Pada Analisis Diskriminan

Populasi	Pengamatan	X_1	X_2	X_3	$\cdots$	X_k
Y_1	1	x_{111}	x_{112}	x_{113}	$\cdots$	x_{11k}
	2	x_{121}	x_{122}	x_{123}	$\cdots$	x_{12k}
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$
	j_1	x_{1j_11}	x_{1j_12}	x_{1j_13}	$\cdots$	x_{1j_1k}
Y_2	1	x_{211}	x_{212}	x_{213}	$\cdots$	x_{21k}
	2	x_{221}	x_{222}	x_{223}	$\cdots$	x_{22k}
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$
	j_2	x_{2j_21}	x_{2j_22}	x_{2j_23}	$\cdots$	x_{2j_2k}
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$
	j_i	x_{ij_i1}	x_{ij_i2}	x_{ij_i3}	$\cdots$	x_{ij_ik}

3.3 Metode Penelitian

Metode yang digunakan dalam penulisan skripsi ini adalah studi pustaka, yaitu dengan mempelajari buku-buku teks penunjang yang berhubungan dengan tugas akhir ini. Kemudian digunakan *software* R dalam pengujian asumsi dan analisis data.

Dalam penelitian ini, langkah-langkah yang dilakukan adalah sebagai berikut.

1. Mencari fungsi diskriminan menggunakan *likelihood rasio test*.
2. Membentuk klasifikasi fungsi analisis diskriminan linear.

3. Mencari total peluang kesalahan klasifikasi /*total probability misclasification* pada analisis fungsi diskriminan untuk dua kelompok.
4. Menduga parameter distribusi normal multivariat dengan menggunakan metode penduga likelihood maksimum dengan langkah-langkah sebagai berikut.
   a. Membentuk fungsi likelihood yang berasal dari fungsi kepekatan peluang distribusi normal multivariat.
   b. Memaksimumkan fungsi yang diperoleh untuk mendapatkan dugaan parameter.
   c. Dugaan yang diperoleh dari metode penduga kemungkinan maksimum diperoleh dengan mencari turunan pertama dari logaritma fungsi kepekatan peluang terhadap parameter-parameter yang hendak diduga dan menyamakannya dengan nol.
5. Menguji asumsi analisis diskriminan, yaitu uji distribusi normal multivariat dan kehomogenan matriks ragam peragam seluruh kelompok dengan menggunakan *software* R.
6. Menerapkan pada data.
   a. Mencari nilai dugaan parameter dengan menghitung nilai rata-rata data $\bar{x}_i$ dan matriks ragam peragam S_i. Nilai rata-rata dan matriks ragam peragam diperoleh dengan menggunakan persamaan:

      $\bar{x}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij}$ dan $S_i = \frac{1}{n_i-1}\left(\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)(x_{ij} - \bar{x}_i)\right)$
   b. Membentuk model fungsi diskriminan linear berdasarkan contoh data dengan jumlah data tiap kelompok yaitu $n_1 = n_2$= 21, sehingga $p_1 < p_2$.

c. Mengklasifikasi data menggunakan aturan klasifikasi analisis diskriminan linear.

d. Menghitung *total probability of misclassification* dua kelompok.

e. Melakukan *resampling* pada data awal untuk memperoleh jumlah data kelompok $n_1 < n_2$ sehingga $p_1 < p_2$.

f. Mencari nilai dugaan parameter dengan menghitung nilai rata-rata data $\bar{x}_i$ dan matriks ragam peragam S_i. Nilai rata-rata dan matriks ragam peragam diperoleh dengan menggunakan persamaan:

$\bar{x}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij}$ dan $S_i = \frac{1}{n_i-1}\left(\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)(x_{ij} - \bar{x}_i)\right)$

g. Membentuk model fungsi diskriminan linear berdasarkan contoh data dengan jumlah data tiap kelompok yaitu $n_1 < n_2$.

h. Mengklasifikasi data menggunakan aturan klasifikasi analisis diskriminan linear.

i. Menghitung *total probability of misclassification* dua kelompok.

j. Melakukan *resampling* pada data awal untuk memperoleh jumlah data kelompok $n_1 > n_2$ sehingga $p_1 > p_2$.

k. Mencari nilai dugaan parameter dengan menghitung nilai rata-rata data $\bar{x}_i$ dan matriks ragam peragam S_i. Nilai rata-rata dan matriks ragam peragam diperoleh dengan menggunakan persamaan:

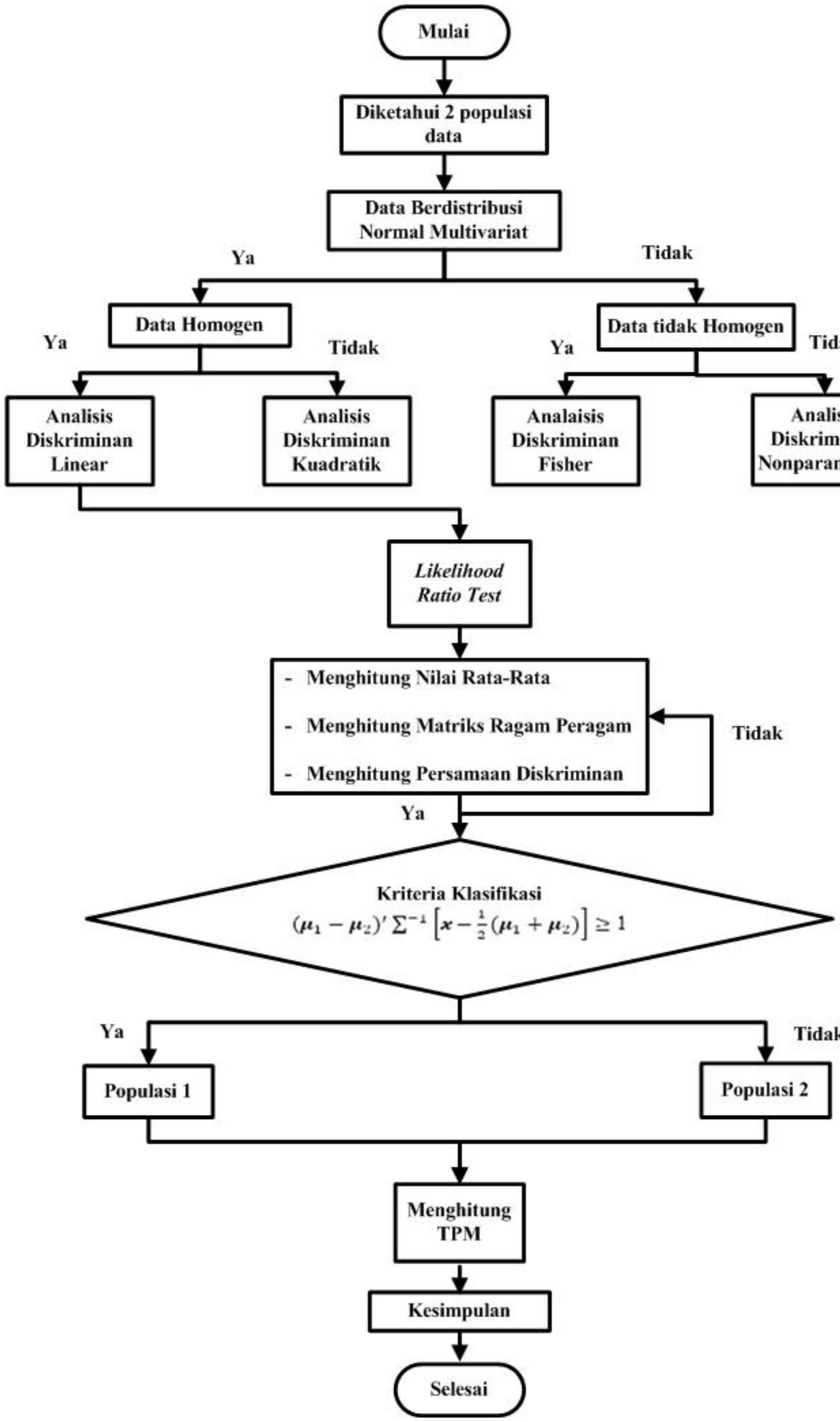
$\bar{x}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij}$ dan $S_i = \frac{1}{n_i-1}\left(\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)(x_{ij} - \bar{x}_i)\right)$

l. Membentuk model fungsi diskriminan linear berdasarkan data contoh data dengan jumlah data tiap kelompok yaitu $n_1 > n_2$.

m. Mengklasifikasi data menggunakan aturan klasifikasi analisis diskriminan linear.

n. Menghitung *total probability of misclassification* dua kelompok.

Secara garis besar langkah-langkah penelitian yang dilakukan dapat tersaji dalam Gambar 3.1.



Gambar 3.1 Diagram Alir Pengklasifikasian Data Dua Kelompok menggunakan Analisis Diskriminan Linear.

V. KESIMPULAN

Dari hasil analisis dan pembahasan, maka dapat diambil kesimpulan sebagai berikut:

1. Analisis diskriminan linear menggunakan *Likelihood Ratio Test* diperoleh dengan membandingkan nilai maksimum fungsi *likelihood*nya.
2. *Total Probality of Misclassification* (TPM) atau total peluang kesalahan klasifikasi suatu pengelompokkan akan minimum jika nilai prior dipertimbangkan ($p_1 \neq p_2$).
3. Berdasarkan contoh data, diperoleh aturan klasifikasi sebagai berikut:

   a. Untuk $p_1 = p_2$

   $$L(\boldsymbol{x}) = [0.183139 \quad 0.472478 \quad 0.147525]\, \boldsymbol{x} - \begin{bmatrix} 51.354762 \\ 3.928571 \\ 50.119048 \end{bmatrix}$$

   b. Untuk $p_1 < p_2$

   $$L(\boldsymbol{x}) = [0.346898 \quad 0.948401 \quad 0.238064]\, \boldsymbol{x} - \begin{bmatrix} 51.326190 \\ 4.361905 \\ 50.395238 \end{bmatrix}$$

   c. Untuk $p_1 > p_2$

   $$L(\boldsymbol{x}) = [0.239803 \quad 0.593401 \quad 0.109227]\, \boldsymbol{x} - \begin{bmatrix} 50.295238 \\ 3.670635 \\ 52.662698 \end{bmatrix}$$

DAFTAR PUSTAKA

Alva, J.A.V., and Estrada, E.G. 2009. *A Generalization of Shapiro Wilk's Test for Multivariate Normality*. Taylor and Francis, Mexico.

Ansori, A. Mattjik dan Made, I Sumertajaya. 2011. *Sidik Peubah Ganda dengan menggunakan SAS*. IPB PRESS, Bogor.

Anton, Howard dan Chris Romes. 2003. *Aljabar Linear Elementer*, Edisi Kelima. (Alih bahasa: Irzam Harmein, Julian Gressando, editor Amalia Safitri). Erlangga, Jakarta.

C., Nayan Giri. 2004. *Multivariate Statistical Analysis Second Edition, Revised and Expanded.* University of Monstreal, Monstreal, Quebec, Canada.

Everitt, B.,S. 2005. *An Rand S-PLUS Companion to Multivariate Analysis*. Spinger, London.

Johnson, R.A., dan Wichern, D.W. 2002. *Applied Multivariate Statistical Analisis*, Fifth Edition. Prentice-Hall, Inc., New Jersey.

Purwo Susongko dan Sari, Inayah Adi, 2012. Analisis Diskriminan dalam Penelitian Ekonomi. Jurnal. Universitas Pancasakti Tegal, Tegal.

Raycov, T., dan Marcoulides, G. 2008. A.An *Introduction to Applied Multivariate Analysis*. Taylor and Fracis Group, New York.

Rencher, A., C. 2002. *Methods of Multivariate Analysis,* Second Edition. John Wiley and Sons, Inc., New York.

S. Srivastava, S., dan M. Caster. E. 1983. *An Introduction to Applied Multivariate*

Statistics. Elsevier Science Publishing Co., Inc., New York.

Sartono, B. dkk. 2003. *Analisis Peubah Ganda*. Institut Pertanian Bogor, Bogor.

Seber, G.A.F., 1983. *Multivariate Observations*. John Wiley and Sons, Inc., New York.

Widiarti, 2003. Landasan Teori Fungsi Diskriminan dan Aplikasinya dengan Matlab. Skripsi. Jurusan Matematika FMIPA UNILA, Bandar Lampung.