

**PERBANDINGAN ANALISIS REGRESI LOGISTIK BINER  
DENGAN ANALISIS DISKRIMINAN PADA DATA TINGKAT  
PENGANGGURAN TERBUKA PROVINSI JAWA BARAT  
DAN JAWA TENGAH TAHUN 2020**

**(Skripsi)**

**Oleh**

**LAURENSIA KLARISA KURI PRAMAYSTI**



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS LAMPUNG  
BANDAR LAMPUNG  
2022**

## **ABSTRACT**

### **COMPARISON OF BINARY LOGISTIC REGRESSION ANALYSIS WITH DISCRIMINANT ANALYSIS OF OPEN UNEMPLOYMENT RATE DATA IN PROVINCE OF WEST JAVA AND CENTRAL JAVA IN 2020**

**By**

**LAURENSIA KLARISA KURI PRAMAYSTI**

Binary logistic regression analysis and discriminant analysis are statistical techniques used to analyze the dichotomous or binary dependent variable. The goal of this study was to compare the two methods based on independent variables that have a significant effect and the level of accuracy on partitioned data with a proportion of 75%-25% and unpartitioned data. This study uses data on the level of open unemployment as the dependent variable and data on the level of labor force participation, the human development index, and the number of dependence as independent variables. The results of this study indicate that the independent variable that has a significant effect on the partitioned data in the binary logistic regression analysis is the labor force participation rate variable, while in the discriminant analysis is the labor force participation rate variable and the human development index and on the data that is not partitioned in the binary logistic regression analysis and discriminant analysis are the same, namely the variables of the labor force participation rate and the human development index. And the level of accuracy on partitioned data (average of training data and validation data) for binary logistic regression analysis is 67.52% while for discriminant analysis is 68.58% and on unpartitioned data for binary logistic regression analysis and discriminant analysis is the same that is equal to 67.74%.

**Keywords :** Binary logistic regression, discriminant, level of accuracy.

## **ABSTRAK**

### **PERBANDINGAN ANALISIS REGRESI LOGISTIK BINER DENGAN ANALISIS DISKRIMINAN PADA DATA TINGKAT PENGANGGURAN TERBUKA PROVINSI JAWA BARAT DAN JAWA TENGAH TAHUN 2020**

**Oleh**

**LAURENSIA KLARISA KURI PRAMAYSTI**

Analisis regresi logistik biner dan analisis diskriminan merupakan teknik statistika yang digunakan untuk menganalisis variabel dependen yang bersifat dikotomis atau biner. Penelitian ini memiliki tujuan untuk membandingkan kedua metode tersebut berdasarkan variabel independen yang berpengaruh signifikan dan tingkat akurasi pada data yang dipartisi dengan proporsi 75%-25% dan data yang tidak dipartisi. Penelitian ini menggunakan data tingkat pengangguran terbuka sebagai variabel dependen dan data tingkat partisipasi angkatan kerja, indeks pembangunan manusia, serta angka ketergantungan sebagai variabel independen. Hasil yang diperoleh dari penelitian ini menunjukkan bahwa variabel independen yang berpengaruh signifikan pada data yang dipartisi dalam analisis regresi logistik biner adalah variabel tingkat partisipasi angkatan kerja sedangkan dalam analisis diskriminan adalah variabel tingkat partisipasi angkatan kerja dan indeks pembangunan manusia dan pada data yang tidak dipartisi dalam analisis regresi logistik biner dan analisis diskriminan adalah sama yaitu variabel tingkat partisipasi angkatan kerja dan indeks pembangunan manusia. Serta tingkat akurasi pada data yang dipartisi (rata-rata data *training* dan data validasi) untuk analisis regresi logistik biner sebesar 67,52% sedangkan untuk analisis diskriminan sebesar 68,58% dan pada data yang tidak dipartisi untuk analisis regresi logistik biner dan analisis diskriminan adalah sama yaitu sebesar 67,74%.

**Kata kunci :** Regresi logistik biner, diskriminan, tingkat akurasi.

**PERBANDINGAN ANALISIS REGRESI LOGISTIK BINER  
DENGAN ANALISIS DISKRIMINAN PADA DATA TINGKAT  
PENGANGGURAN TERBUKA PROVINSI JAWA BARAT  
DAN JAWA TENGAH TAHUN 2020**

**Oleh**

**LAURENSIA KLARISA KURI PRAMAYSTI**

**Skripsi**

**Sebagai Salah Satu Syarat untuk Memperoleh Gelar  
SARJANA MATEMATIKA**

**Pada**

**Jurusan Matematika  
Fakultas Matematika dan Ilmu Pengetahuan Alam  
Universitas Lampung**



**JURUSAN MATEMATIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS LAMPUNG  
BANDAR LAMPUNG  
2022**

Judul Skripsi

: **PERBANDINGAN ANALISIS REGRESI LOGISTIK  
BINER DENGAN ANALISIS DISKRIMINAN  
PADA DATA TINGKAT PENGANGGURAN  
TERBUKA PROVINSI JAWA BARAT DAN  
JAWA TENGAH TAHUN 2020**

Nama Mahasiswa

: **Laurensia Klarisa Kuri Pramaysti**

Nomor Pokok Mahasiswa : **1817031068**

Jurusan

: **Matematika**

Fakultas

: **Matematika dan Ilmu Pengetahuan Alam**



1. **Komisi Pembimbing**

  
**Prof. Drs. Mustofa Usman, M.A., Ph.D.**  
NIP 19570101 198403 1 020

  
**Dra. Dorrah Aziz, M.Si.**  
NIP 19610128 198811 2 001

2. **Ketua Jurusan Matematika**

  
**Dr. Aang Nuryaman, S.Si., M.Si.**  
NIP 19740316 200501 1 001

**MENGESAHKAN**

**1. Tim Penguji**

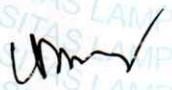
**Ketua : Prof. Drs. Mustofa Usman, M.A., Ph.D.** .....



**Sekretaris : Dra. Dorrah Aziz, M.Si.** .....



**Penguji  
Bukan Pembimbing : Ir. Warsono, M.S., Ph.D.** .....



**2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam**



**Dr. Eng. Sripto Dwi Yuwono, S.Si., M.T.**  
NIP. 19740705 200003 1 001



**Tanggal Lulus Ujian Skripsi : 19 Juli 2022**

## PERNYATAAN SKRIPSI MAHASISWA

Saya yang bertanda tangan di bawah ini:

Nama : Laurensia Klarisa Kuri Pramaysti

Nomor Pokok Mahasiswa : 1817031068

Jurusan : Matematika

Judul Skripsi : Perbandingan Analisis Regresi Logistik Biner dengan Analisis Diskriminan Pada Data Tingkat Pengangguran Terbuka Provinsi Jawa Barat dan Jawa Tengah Tahun 2020

Dengan ini menyatakan bahwa skripsi ini adalah hasil pekerjaan saya sendiri, bukan hasil orang lain dan semua hasil tulisan yang tertuang dalam skripsi ini telah mengikuti kaidah penulisan karya ilmiah Universitas Lampung. Apabila kemudian hari terbukti bahwa skripsi ini merupakan hasil salinan atau dibuat oleh orang lain, maka saya bersedia menerima sanksi sesuai dengan ketentuan akademik yang berlaku.

Bandar Lampung, 19 Juli 2022

Penulis



Laurensia Klarisa Kuri Pramaysti  
NPM. 1817031068

## **RIWAYAT HIDUP**

Penulis dilahirkan di Kota Tangerang, Banten pada 13 April 2000, dari pasangan Bapak Ladislaus Sundakir dan Ibu Sri Handayani sebagai anak pertama dari dua bersaudara.

Penulis menempuh pendidikan dimulai dari SD Negeri Gandasari 2 yang selesai pada tahun 2012, lalu dilanjutkan ke SMP Negeri 2 Jawilan yang selesai pada tahun 2015, kemudian diteruskan ke SMA Negeri 1 Cikande yang selesai pada tahun 2018, hingga akhirnya terdaftar sebagai mahasiswa S1 Jurusan Matematika, Fakultas Ilmu Pengetahuan Alam, Universitas Lampung melalui jalur masuk SBMPTN pada tahun 2018.

Selama menjadi mahasiswa, penulis aktif dalam sebuah organisasi yaitu Himpunan Mahasiswa Jurusan Matematika (HIMATIKA) sebagai anggota bidang Keilmuan periode 2019 dan 2020.

Pada awal tahun 2021, penulis melaksanakan Kuliah Kerja Nyata (KKN) di Desa Junti, Kecamatan Jawilan, Kabupaten Serang, Banten. Kemudian, pada pertengahan tahun 2021 penulis melaksanakan Kerja Praktik di Rumah Sakit Ibu dan Anak Santa Anna Bandar Lampung.

## **PERSEMBAHAN**

Puji dan syukur saya haturkan kepada Tuhan Yesus, atas segala berkat, karunia dan limpahan-Nya sehingga sebuah karya sederhana namun istimewa ini dapat terselesaikan. Karya ini saya persembahkan kepada :

Orang tua saya terkasih, Bapak Ladislaus Sundakir dan Ibu Sri Handayani atas kesempatan dan kepercayaannya pada saya serta segenap doa, kasih sayang, dan perjuangan yang tidak dapat diukur oleh apapun.

Adik tersayang, Vincensius Galih Raditya yang selalu memberikan keceriaan dan semangat.

Bapak, Ibu dosen yang telah sangat baik membagikan banyak ilmu luar biasa dan membantu selama perkuliahan.

Teman-teman terbaik yang selalu mendoakan, mendukung, membantu dan memberikan semangat.

Almamater tercinta, Universitas Lampung.

## **KATA INSPIRASI**

Ia membuat segala sesuatu indah pada waktunya.

(Pengkhotbah 3 : 11)

Serahkanlah perbuatanmu kepada Tuhan, maka terlaksanalah segala rencanamu.

(Amsal 16 : 3)

Apa saja yang kamu minta dalam doa dengan penuh kepercayaan, kamu akan menerimanya.

(Matius 21 : 22)

Bukan karena hari ini indah kita bahagia, tetapi karena kita bahagia maka hari ini menjadi indah. Bukan karena mudah kita yakin bisa, tetapi karena kita yakin bisa semua menjadi mudah.

(W.S Rendra)

Tidak masalah kita kalah satu-dua pertarungan, atau malah kalah berkali-kali, tetapi pastikan kitalah yang tetap berdiri tegak diakhir semua kisah.

(Tere Liye, novel "Komet Minor")

## SANWACANA

Puji syukur kehadiran Tuhan Yang Maha Esa karena atas berkat limpahan karunia dan izin-ya sehingga penulis dapat menyelesaikan skripsi dengan judul “Perbandingan Analisis Regresi Logistik Biner dengan Analisis Diskriminan Pada Data Tingkat Pengangguran Terbuka Provinsi Jawa Barat dan Jawa Tengah Tahun 2020” pada waktu yang tepat sebagai salah satu syarat untuk mencapai gelar Sarjana Matematika pada program S1 Matematika Jurusan Matematika Fakultas MIPA Universitas Lampung.

Dalam menyelesaikan skripsi ini, penulis mendapat banyak bantuan, bimbingan dan dukungan dari berbagai pihak. Oleh karena itu, penulis ucapkan terima kasih kepada:

1. Bapak Prof. Drs. Mustofa Usman, M.A., Ph.D. selaku dosen pembimbing pertama yang telah memberikan arahan, masukan, saran dan motivasi serta meluangkan waktu ditengah kesibukan selama penyusunan skripsi ini.
2. Ibu Dra. Dorrah Aziz, M.Si. selaku dosen pembimbing kedua yang telah memberikan kritik, saran dan motivasi serta meluangkan waktu ditengah kesibukan selama penyusunan skripsi ini.
3. Bapak Ir. Warsono, M.S., Ph.D. selaku dosen pembahas yang telah memberikan kritik dan saran yang membangun dalam penyelesaian skripsi ini.
4. Bapak Dr. Muslim Ansori, S.Si., M.Si. selaku dosen pembimbing akademik yang telah membantu dan memberikan arahan selama masa perkuliahan.
5. Bapak Dr. Aang Nuryaman, S.Si., M.Si. selaku Ketua Jurusan Matematika FMIPA Universitas Lampung.
6. Bapak Dr. Eng. Suropto Dwi Yuwono, S.Si., M.T. selaku Dekan FMIPA Universitas Lampung.

7. Seluruh Dosen dan Karyawan Jurusan Matematika FMIPA Universitas Lampung.
8. Bapak Ladislaus Sundakir dan Ibu Sri Handayani terkasih yang selalu mendukung, mendoakan, berkorban dan bersabar hingga skripsi ini terselesaikan.
9. Adik Vincensius Galih Raditya tersayang yang selalu berbagi keceriaan dan semangat kepada penulis selama penyelesaian skripsi ini.
10. Teman-teman terbaik Riska, Ria, Spalda, Yosi, Nurma, Verdi, Mita, Mona, Novita, Lutfia yang selalu berbagi semangat dan keceriaan serta selalu mendukung penulis selama penyelesaian skripsi ini.
11. Teman-teman seperbimbingan yang selalu berjuang bersama dalam penyelesaian skripsi ini.
12. Teman-teman seperjuangan Matematika 2018 yang tidak bisa penulis tuliskan satu-persatu.
13. Seluruh pihak yang telah terlibat dan membantu dalam penyelesaian skripsi ini yang belum penulis sebutkan.

Akhir kata, semoga ketulusan serta bantuan dari semua pihak yang terlibat dalam penyelesaian skripsi ini mendapatkan berkat dan balasan baik dari Tuhan Yang Maha Esa.

Bandar Lampung, 19 Juli 2022  
Penulis,

Laurensia Klarisa Kuri Pramaysti

## DAFTAR ISI

	Halaman
<b>DAFTAR TABEL .....</b>	<b>viii</b>
<b>I. PENDAHULUAN .....</b>	<b>1</b>
1.1 Latar Belakang dan Masalah .....	1
1.2 Tujuan Penelitian .....	3
1.3 Manfaat Penelitian .....	3
<b>II. TINJAUAN PUSTAKA .....</b>	<b>4</b>
2.1 Analisis Multivariat .....	4
2.2 Partisi Sampel .....	4
2.3 Analisis Regresi Logistik .....	5
2.4 Analisis Regresi Logistik Biner .....	5
2.5 <i>Maximum Likelihood Estimation</i> .....	7
2.6 Pengujian Parameter Secara Simultan .....	11
2.7 Pengujian Parameter Secara Parsial .....	11
2.8 Pengujian Kesesuaian Model .....	12
2.9 Interpretasi Koefisien Model .....	13
2.10 Analisis Diskriminan .....	13
2.10.1 Fungsi Diskriminan Dua Kelompok .....	14
2.10.2 Pengujian Perbedaan Rata-rata Kelompok .....	15
2.11 Metode Untuk Membentuk Fungsi Diskriminan .....	16
2.12 Pengujian Normal Multivariat .....	17
2.13 Pengujian Kesamaan Matriks Kovarian .....	18
2.14 Estimasi Ketepatan Klasifikasi .....	19
<b>III. METODOLOGI PENELITIAN .....</b>	<b>22</b>
3.1 Waktu dan Tempat Penelitian .....	22
3.2 Data Penelitian .....	22
3.3 Metode Penelitian .....	23
<b>IV. HASIL DAN PEMBAHASAN .....</b>	<b>25</b>
4.1 Statistika Deskriptif .....	25
4.2 Pengujian Asumsi .....	27
4.2.1 Pengujian Asumsi Normal Multivariat .....	27
4.2.2 Pengujian Asumsi Kesamaan Matriks Kovarian .....	28
4.3 Analisis Regresi Logistik Biner .....	29

4.3.1	Model Awal Regresi Logistik Biner .....	29
4.3.2	Pengujian Signifikansi Parameter Secara Simultan .....	30
4.3.3	Pengujian Signifikansi Parameter Secara Parsial .....	32
4.3.4	Model Baru Regresi Logistik Biner .....	33
4.3.5	Pengujian Kesesuaian Model.....	36
4.3.6	Model Akhir Regresi Logistik Biner.....	37
4.3.7	Menghitung Nilai APER dan Tingkat Akurasi.....	38
4.4	Analisis Diskriminan.....	41
4.4.1	Pengujian Kesamaan Vektor Rata-rata Kelompok .....	41
4.4.2	Pengujian Signifikansi Variabel .....	42
4.4.3	Membentuk Fungsi Diskriminan .....	44
4.4.4	Menghitung Nilai APER dan Tingkat Akurasi.....	45
4.5	Membandingkan Nilai APER dan Tingkat Akurasi dari Metode Regresi .....	48
	Logistik Biner dengan Metode Diskriminan .....	48
<b>V.</b>	<b>KESIMPULAN.....</b>	<b>49</b>
	<b>DAFTAR PUSTAKA.....</b>	<b>51</b>
	<b>LAMPIRAN .....</b>	<b>53</b>

## DAFTAR TABEL

Tabel	Halaman
1. Matriks Konfusi Hasil Klasifikasi .....	20
2. Pengelompokkan Variabel Dependen .....	25
3. Pengelompokkan Variabel Independen .....	26
4. Hasil Pengujian Normal Multivariat .....	28
5. Hasil Pengujian Box M .....	28
6. Hasil Pendugaan Model Data <i>Training</i> .....	29
7. Hasil Pendugaan Model Data Keseluruhan .....	30
8. Statistik Uji <i>Ratio Likelihood</i> Data <i>Training</i> .....	31
9. Statistik Uji <i>Ratio Likelihood</i> Data Keseluruhan .....	32
10. Statistik Uji Wald Data <i>Training</i> .....	32
11. Statistik Uji Wald Data Keseluruhan .....	33
12. Hasil Pendugaan Model Baru Data <i>Training</i> .....	34
13. Statistik Uji <i>Ratio Likelihood</i> Model Baru Data <i>Training</i> .....	34
14. Hasil Pendugaan Model Baru Data Keseluruhan .....	35
15. Statistik Uji <i>Ratio Likelihood</i> Model Baru Data Keseluruhan .....	35
16. Statistik Hosmer-Lemeshow <i>goodness of fit</i> Data <i>Training</i> .....	36
17. Statistik Hosmer-Lemeshow <i>goodness of fit</i> Data Keseluruhan .....	37
18. Matriks Konfusi Data <i>Training</i> dari Data Partisi .....	39
19. Matriks Konfusi Data Validasi dari Data Partisi .....	39
20. Matriks Konfusi Data Keseluruhan .....	40

21. Statistik $T^2$ Hotelling Data Training .....	41
22. Statistik $T^2$ Hotelling Data Keseluruhan .....	42
23. Hasil Metode Seleksi <i>Stepwise</i> Data Training .....	43
24. Hasil Metode Seleksi <i>Stepwise</i> Data Keseluruhan .....	43
25. Fungsi Diskriminan Data Training .....	44
26. Fungsi Diskriminan Data Keseluruhan .....	45
27. Matriks Konfusi Data Training dari Data Partisi .....	46
28. Matriks Konfusi Data Validasi dari Data Partisi .....	46
29. Matriks Konfusi Data Keseluruhan .....	47
30. Nilai APER dan Tingkat Akurasi Metode Regresi Logistik Biner dan Metode Diskriminan .....	48

## I. PENDAHULUAN

### 1.1 Latar Belakang dan Masalah

Analisis multivariat adalah metode statistika yang secara simultan melakukan analisis terhadap lebih dari dua variabel pada setiap objek atau orang. Analisis multivariat terbagi menjadi dua bagian menurut hubungan antar variabelnya, yaitu interdependensi dan dependensi. Analisis multivariat interdependensi didefinisikan memuat variabel-variabel yang tidak saling bergantung satu sama lain dan analisis multivariat dependensi didefinisikan memuat variabel-variabel yang saling bergantung satu sama lain (Santoso, 2017).

Dalam analisis multivariat dependensi, teknik statistika yang dapat digunakan untuk menganalisis variabel dependen yang bersifat dikotomus atau biner salah satunya yaitu analisis diskriminan. Analisis diskriminan adalah analisis multivariat yang digunakan untuk memperkirakan hubungan antar variabel dependen yang bersifat kategorik (dikotomus atau multikotomus) dengan variabel independen yang bersifat kuantitatif. Tujuan dari analisis diskriminan adalah membangun satu set diskriminan yang digunakan untuk menggambarkan pemisahan kelompok berdasarkan sekumpulan variabel yang direduksi, menganalisis kontribusi variabel asli terhadap pemisahan, dan mengevaluasi tingkat pemisahan (Timm, 2002).

Selain analisis diskriminan, terdapat juga teknik statistika yang dapat digunakan untuk menganalisis variabel dependen yang bersifat dikotomus atau biner yaitu analisis regresi logistik. Analisis regresi logistik adalah analisis yang menggambarkan hubungan antara variabel dependen yang bersifat kategorik

(biner, multinomial maupun ordinal) dan satu set variabel independen yang bersifat kuantitatif ataupun kualitatif. Tujuan dari analisis regresi logistik adalah untuk memperkirakan probabilitas memilih masing-masing variabel dependen dan untuk menyatakan hasil dari variabel independen dalam *odds ratio* (Hosmer, Lemeshow, & Studirvant, 2013).

Berdasarkan kedua teknik statistika yang dapat digunakan untuk menganalisis variabel dependen yang bersifat dikotomis atau biner tersebut, maka sangat menarik untuk mengkaji perbandingan hasilnya berdasarkan variabel independen yang berpengaruh signifikan terhadap model dan tingkat akurasi yang lebih baik. Namun, dalam penelitian ini dikhususkan untuk data Tingkat Pengangguran Terbuka Provinsi Jawa Barat dan Jawa Tengah tahun 2020.

Penelitian dengan mengkaji perbandingan analisis diskriminan dan analisis regresi logistik pernah dilakukan oleh Mohammed M. A. A. M dalam jurnalnya pada tahun 2018 dengan menggunakan data Indeks Pembangunan Manusia pada tahun 2016 dari 96 negara berdasarkan faktor-faktor yang mempengaruhinya. Pada penelitian tersebut data yang digunakan tidak dipartisi menjadi data *training* dan data validasi serta diketahui bahwa metode regresi logistik lebih efisien dibandingkan dengan metode diskriminasi dalam akurasi prediksi. Dan juga pernah dilakukan oleh Wd. Rifqah, Resmawan dan Ismail Djakaria dalam jurnalnya pada tahun 2019 dengan menggunakan data jurusan yang dapat dipilih siswa kelas X Madrasah Aliyah Negeri Model Gorontalo berdasarkan faktor-faktor yang mempengaruhinya. Pada penelitian tersebut analisis regresi logistik yang digunakan yaitu analisis regresi logistik multinomial dan data yang digunakan tidak dipartisi menjadi data *training* dan data validasi serta diketahui bahwa kedua metode tersebut sama baiknya dengan tingkat ketepatan klasifikasi sebesar 53,60%.

Dalam penelitian ini, hal yang membedakan dengan penelitian sebelumnya tersebut, yaitu data akan dianalisis secara keseluruhan dan juga dilakukan partisi menjadi data *training* dan data validasi.

## **1.2 Tujuan Penelitian**

Adapun tujuan dari penelitian ini adalah :

1. Membandingkan tingkat akurasi metode regresi logistik biner dengan metode diskriminan dalam mengklasifikasikan tingkat pengangguran terbuka di provinsi Jawa Barat dan Jawa Tengah tahun 2020 berdasarkan data yang dipartisi dan tidak dipartisi.
2. Menganalisis variabel independen yang berpengaruh signifikan terhadap model tingkat pengangguran terbuka di provinsi Jawa Barat dan Jawa Tengah tahun 2020 dengan metode regresi logistik biner dan metode diskriminan berdasarkan data yang dipartisi dan tidak dipartisi.

## **1.3 Manfaat Penelitian**

Adapun manfaat dari penelitian ini adalah :

1. Menambah pengetahuan dalam penerapan analisis regresi logistik biner dan analisis diskriminan.
2. Memperoleh hasil perbandingan tingkat akurasi metode regresi logistik biner dengan metode diskriminan.
3. Mengetahui variabel independen yang berpengaruh signifikan terhadap model tingkat pengangguran terbuka di provinsi Jawa Barat dan Jawa Tengah tahun 2020.

## II. TINJAUAN PUSTAKA

### 2.1 Analisis Multivariat

Analisis multivariat adalah metode statistik yang melakukan analisis secara bersama-sama (simultan) terhadap dua variabel atau lebih pada setiap objek atau orang. Dalam analisis multivariat terdapat dua klasifikasi yang membagi berbagai teknik multivariat berdasarkan hubungan antar variabelnya, yaitu klasifikasi interdependensi dan klasifikasi dependensi (Santoso, 2017).

Klasifikasi interdependensi didefinisikan memuat variabel-variabel yang tidak saling bergantung satu sama lain. Teknik multivariat yang termasuk dalam klasifikasi interdependensi antara lain analisis faktor, analisis kluster, MDS (Multidimensional Scaling), dan CA (Categorical Analysis). Sedangkan klasifikasi dependensi didefinisikan memuat variabel-variabel yang saling bergantung satu sama lain. Teknik multivariat yang termasuk dalam klasifikasi dependensi antara lain analisis regresi berganda, analisis regresi logistik, analisis diskriminan, SEM, MANOVA, dan analisis korelasi kanonikal (Santoso, 2017).

### 2.2 Partisi Sampel

Salah satu cara untuk menjauhi bias adalah dengan membagi sampel menjadi dua bagian, yaitu *training sample* dan *validation sample*. *Training sample* digunakan untuk membangun fungsi klasifikasi sedangkan *validation sample* yang digunakan untuk mengevaluasinya (Rencher, 2002). Tidak ada pedoman pasti yang telah ditetapkan untuk menentukan ukuran relatif dari *training sample* dan *validation*

*sample*. Pendekatan yang paling populer adalah membagi total sampel secara acak sehingga setengah dari sampel ditempatkan dalam *training sample* dan setengah lainnya ditempatkan dalam *validation sample*. Beberapa peneliti lebih memilih pemisahan 60%–40% atau bahkan 75%–25% antara *training sample* dan *validation sample*, tergantung pada ukuran sampel keseluruhan (Hair, et al., 2019) dan menurut Rencher (2002), lebih suka menggunakan semua atau hampir semua data untuk membangun fungsi klasifikasi untuk meminimalkan varians dari tingkat kesalahan.

### **2.3 Analisis Regresi Logistik**

Regresi logistik adalah suatu model yang mendeskripsikan hubungan antara variabel dependen dan satu set variabel independen (Hosmer, Lemeshow, & Studirvant, 2013). Regresi logistik dapat dibedakan menjadi tiga, yaitu regresi logistik respon biner, ordinal dan multinomial. Regresi logistik biner dan multinomial adalah ketika variabel dependen atau responnya berskala nominal, jika biner berarti terdiri dari dua kategori dan jika multinomial terdiri dari tiga atau lebih kategori. Sedangkan regresi logistik ordinal adalah ketika variabel dependennya berskala ordinal seperti contoh pada prestasi kerja (tidak memadai, memuaskan, luar biasa). Tujuan dari regresi logistik yaitu untuk memperkirakan probabilitas memilih masing-masing variabel dependen dan untuk menyatakan hasil dari variabel independen dalam *odds ratio*.

### **2.4 Analisis Regresi Logistik Biner**

Regresi logistik biner adalah metode yang digunakan untuk menganalisis hubungan antara variabel dependen dan satu set variabel independen, dengan variabel dependen bersifat dikotomus atau biner dan variabel independennya dengan skala pengukuran yang berbeda (Hosmer, Lemeshow, & Studirvant, 2013). Karena variabel dependen ( $y$ ) bersifat biner yang artinya terdiri dari dua

kelompok, maka dimisalkan dengan  $Y = 1$  dan  $Y = 0$ . Probabilitas bersyarat untuk  $Y = 1$  dinotasikan dengan:

$$\Pr(Y = 1|\mathbf{x}) = \pi(\mathbf{x}) \quad (2.1)$$

dan probabilitas untuk  $Y = 0$  dinotasikan dengan:

$$\Pr(Y = 0|\mathbf{x}) = 1 - \pi(\mathbf{x}). \quad (2.2)$$

Model regresi logistik yang mencakup dua atau lebih variabel independen dapat disebut sebagai multivariabel atau model regresi logistik multipel. Model regresi logistik multipel dengan  $m$  variabel independen diberikan pada persamaan sebagai berikut.

$$\pi(\mathbf{x}) = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m}} \quad (2.3)$$

Karena fungsi  $\pi(x)$  merupakan fungsi yang tidak linier dan untuk membuat fungsi tersebut menjadi linier maka perlu dilakukan transformasi. Transformasi yang dilakukan yaitu transformasi logit. Fungsi  $\pi(x)$  bila dilakukan transformasi logit yaitu sebagai berikut.

$$\begin{aligned} \pi(\mathbf{x}) &= \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m}} \\ (\pi(\mathbf{x})) (1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m}) &= e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m} \\ (\pi(\mathbf{x})) + (\pi(\mathbf{x}) e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m}) &= e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m} \\ \pi(\mathbf{x}) &= e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m} - (\pi(\mathbf{x}) e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m}) \\ \pi(\mathbf{x}) &= (1 - \pi(\mathbf{x})) e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m} \\ \frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})} &= e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m} \\ \ln\left(\frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})}\right) &= \ln\left(e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m}\right) \\ \ln\left(\frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})}\right) &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m \end{aligned}$$

Sehingga, transformasi logit dari fungsi  $\pi(\mathbf{x})$  yaitu sebagai berikut.

$$g(\mathbf{x}) = \ln\left(\frac{\pi(\mathbf{x})}{1-\pi(\mathbf{x})}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m \quad (2.4)$$

dengan  $g(\mathbf{x})$  dikatakan sebagai fungsi logit dengan  $m$  variabel independen.

## 2.5 Maximum Likelihood Estimation

Metode *maximum likelihood estimation* merupakan suatu pendugaan untuk mengestimasi nilai parameter yang tidak diketahui. Pada regresi logistik, metode *maximum likelihood estimation* akan memberikan nilai estimasi parameter  $\beta$  dengan cara memaksimalkan fungsi *likelihood* (Hosmer, Lemeshow, & Studirvant, 2013). Secara umum, fungsi probabilitas dari pasangan variabel independen dan dependen pada pengamatan ke- $i$  ( $x_i, y_i$ ) dengan  $i = 1, 2, \dots, n$  yaitu sebagai berikut.

$$f(y_i) = \pi(\mathbf{x}_i)^{y_i} [1 - \pi(\mathbf{x}_i)]^{1-y_i}; Y = 0, 1 \quad (2.5)$$

dengan

$$\pi(\mathbf{x}_i) = \frac{e^{\left(\sum_{j=0}^m \beta_j x_{ij}\right)}}{1 + e^{\left(\sum_{j=0}^m \beta_j x_{ij}\right)}}$$

dimana ketika  $j=0$  maka nilai dari  $x_{ij} = x_{i0} = 1$ . Karena setiap pasangan diasumsikan independen maka fungsi *likelihood* merupakan kumpulan dari masing-masing pasangan fungsi yaitu sebagai berikut.

$$\begin{aligned} l(\boldsymbol{\beta}) &= \prod_{i=1}^n [\pi(\mathbf{x}_i)^{y_i} [1 - \pi(\mathbf{x}_i)]^{1-y_i}] \\ &= \left\{ \prod_{i=1}^n \pi(\mathbf{x}_i)^{y_i} \right\} \left\{ \prod_{i=1}^n ((1 - \pi(\mathbf{x}_i))^{1-y_i}) \right\} \\ &= \left\{ \prod_{i=1}^n (1 - \pi(\mathbf{x}_i)) \right\} \left\{ \prod_{i=1}^n \pi(\mathbf{x}_i)^{y_i} \right\} \left\{ \prod_{i=1}^n (1 - \pi(\mathbf{x}_i))^{-y_i} \right\} \end{aligned}$$

$$\begin{aligned}
&= \left\{ \prod_{i=1}^n (1 - \pi(\mathbf{x}_i)) \right\} \left\{ \prod_{i=1}^n e^{\left( \ln \left( \frac{\pi(\mathbf{x}_i)}{1 - \pi(\mathbf{x}_i)} \right)^{y_i} \right)} \right\} \\
&= \left\{ \prod_{i=1}^n (1 - \pi(\mathbf{x}_i)) \right\} \left\{ e^{\left( \sum_{i=1}^n y_i \ln \left( \frac{\pi(\mathbf{x}_i)}{1 - \pi(\mathbf{x}_i)} \right) \right)} \right\} \\
&= \left\{ \prod_{i=1}^n \left( \frac{1}{1 + e^{\sum_{j=0}^m \beta_j x_{ij}}} \right) \right\} \left\{ e^{\left( \sum_{j=0}^m \beta_j \left( \sum_{i=1}^n y_i x_{ij} \right) \right)} \right\}
\end{aligned}$$

dimana  $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_m)$ . Untuk memudahkan perhitungan matematis, maka dilakukan dengan log *likelihood* yang dinyatakan sebagai berikut.

$$\begin{aligned}
L(\boldsymbol{\beta}) &= \ln(l(\boldsymbol{\beta})) \\
&= \ln \left[ \left\{ \prod_{i=1}^n \left( \frac{1}{1 + e^{\sum_{j=0}^m \beta_j x_{ij}}} \right) \right\} \left\{ e^{\left( \sum_{j=0}^m \beta_j \left( \sum_{i=1}^n y_i x_{ij} \right) \right)} \right\} \right] \\
&= \left\{ \ln \left( e^{\left( \sum_{j=0}^m \beta_j \left( \sum_{i=1}^n y_i x_{ij} \right) \right)} \right) \right\} \left\{ \ln \left( \prod_{i=1}^n \left( 1 + e^{\sum_{j=0}^m \beta_j x_{ij}} \right)^{-1} \right) \right\} \\
&= \sum_{j=0}^m \beta_j \left( \sum_{i=1}^n y_i x_{ij} \right) - \sum_{i=1}^n \ln \left( 1 + e^{\sum_{j=0}^m \beta_j x_{ij}} \right)
\end{aligned}$$

Untuk mencari nilai  $\boldsymbol{\beta}$  dan  $L(\boldsymbol{\beta})$  dilakukan dengan membuat turunan pertama dari  $L(\boldsymbol{\beta})$  terhadap  $\boldsymbol{\beta}$  dan disamakan dengan nol.

$$\begin{aligned}
\frac{\partial L(\boldsymbol{\beta})}{\partial \beta} &= \frac{\partial}{\partial \beta} \left[ \sum_{j=0}^m \beta_j \left( \sum_{i=1}^n y_i x_{ij} \right) - \sum_{i=1}^n \ln \left( 1 + e^{\sum_{j=0}^m \beta_j x_{ij}} \right) \right] = 0 \\
&= \sum_{i=1}^n y_i x_{ij} - \sum_{i=1}^n x_{ij} \left( \frac{e^{\sum_{j=0}^m \beta_j x_{ij}}}{1 + e^{\sum_{j=0}^m \beta_j x_{ij}}} \right) = 0
\end{aligned}$$

sehingga,

$$\begin{aligned}
\sum_{i=1}^n y_i x_{ij} - \sum_{i=1}^n x_{ij} \pi(x_i) &= 0 \\
\sum_{i=1}^n x_{ij} (y_i - \pi(x_i)) &= 0
\end{aligned} \tag{2.6}$$

dengan  $j = 0, 1, 2, \dots, m$ .

Metode untuk memperkirakan varians dan kovarians dari koefisien parameter yang diperkirakan dikembangkan melalui teori *maximum likelihood estimation* yang dikembangkan dengan baik (Rao, 1973 dalam Hosmer, Lemeshow, & Studirvant, 2013). Teori tersebut menyatakan bahwa perkiraan varians dan kovarians diperoleh dari turunan parsial kedua dari fungsi *log-likelihood* sebagai berikut.

$$\frac{\partial^2 L(\beta)}{\partial \beta_j \partial \beta_s} = - \sum_{i=1}^n x_{ij} x_{is} \pi_i (1 - \pi_i) \quad (2.7)$$

dengan  $j, s = 0, 1, 2, \dots, m$ , dimana  $\pi_i$  menunjukkan  $\pi(x_i)$ . Matriks varians dan kovarians dari parameter yang diestimasi diperoleh melalui invers matriks sebagai berikut.

$$\text{Var}(\hat{\beta}) = (\mathbf{X}' \hat{\mathbf{V}} \mathbf{X})^{-1} \quad (2.8)$$

dimana  $\mathbf{X}$  adalah matriks  $n \times (m+1)$  dari setiap pengamatan dan  $\hat{\mathbf{V}}$  adalah matriks diagonal  $n \times n$  dengan diagonal utamanya adalah  $\hat{\pi}_i (1 - \hat{\pi}_i)$ . Matriks  $\mathbf{X}$  yaitu sebagai berikut

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1m} \\ 1 & x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix}$$

dan matriks  $\hat{\mathbf{V}}$  sebagai berikut

$$\hat{\mathbf{V}} = \begin{bmatrix} \hat{\pi}_1 (1 - \hat{\pi}_1) & 0 & 0 & 0 \\ 0 & \hat{\pi}_2 (1 - \hat{\pi}_2) & 0 & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & \cdots & 0 & \hat{\pi}_n (1 - \hat{\pi}_n) \end{bmatrix}$$

Estimasi standar *error* ( $SE(\hat{\beta})$ ) dari koefisien parameter estimasi yaitu akar kuadrat diagonal utama. Untuk mendapatkan nilai estimasi  $\beta$  dari turunan pertama fungsi *log likelihood* yang merupakan non linier maka digunakan metode iterasi Newton Raphson. Persamaan yang digunakan yaitu sebagai berikut.

$$\beta_{(t+1)} = \beta_{(t)} - (\mathbf{H}_{(t)})^{-1} \mathbf{q}_{(t)} \quad (2.9)$$

dimana  $t = 1, 2, 3$ , konvergen. Dengan,

$$\mathbf{q}' = \left( \frac{\partial L(\boldsymbol{\beta})}{\partial \beta_0} \quad \frac{\partial L(\boldsymbol{\beta})}{\partial \beta_1} \quad \dots \quad \frac{\partial L(\boldsymbol{\beta})}{\partial \beta_k} \right)$$

dan  $\mathbf{H}$  merupakan matriks Hessian dari  $\boldsymbol{\beta}$  yaitu sebagai berikut

$$\mathbf{H} = \begin{bmatrix} \frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_0^2} & \frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_0 \partial \beta_1} & \frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_0 \partial \beta_2} & \dots & \frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_0 \partial \beta_m} \\ \frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_0 \partial \beta_1} & \frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_1^2} & \frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_1 \partial \beta_2} & \dots & \frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_1 \partial \beta_m} \\ \frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_0 \partial \beta_2} & \frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_1 \partial \beta_2} & \frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_2^2} & \vdots & \frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_2 \partial \beta_m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_0 \partial \beta_m} & \frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_1 \partial \beta_m} & \frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_2 \partial \beta_m} & \dots & \frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_m^2} \end{bmatrix}$$

dan pada setiap iterasi berlaku

$$\frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_j \partial \beta_s} = - \sum_{i=1}^n x_{ij} x_{is} \pi_{i(t)} (1 - \pi_{i(t)}) \quad (2.10)$$

$$\frac{\partial L(\boldsymbol{\beta})}{\partial \beta_j} = \sum_{i=1}^n (y_i - \pi_{i(t)}) x_{ij} = 0 \quad (2.11)$$

$$\pi_{i(t)} = \frac{e^{(\sum_{j=0}^m \beta_{j(t)} x_{ij})}}{1 + e^{(\sum_{j=0}^m \beta_{j(t)} x_{ij})}} \quad (2.12)$$

Langkah-langkah dalam melakukan iterasi Newton Raphson yaitu sebagai berikut.

- 1) Menentukan nilai awal estimasi  $\boldsymbol{\beta}_{(0)}$  dan selanjutnya menghitung persamaan (2.9) sehingga diperoleh  $\pi_{i(0)}$ .
- 2) Dengan  $\pi_{i(0)}$  yang telah didapatkan diperoleh matriks Hessian  $\mathbf{H}_{(0)}$  dan vektor  $\mathbf{q}_{(0)}$ .
- 3) Untuk  $t > 0$  digunakan persamaan (2.9) dan (2.12) hingga mendapatkan  $\pi_{i(t)}$  dan  $\boldsymbol{\beta}_{(t)}$  yang konvergen.

## 2.6 Pengujian Parameter Secara Simultan

Pengujian parameter secara simultan dilakukan dengan menggunakan uji *Ratio Likelihood* yang bermaksud untuk menguji signifikansi koefisien  $\beta$  secara keseluruhan (Hosmer, Lemeshow, & Studirvant, 2013). Adapun dengan hipotesis sebagai berikut:

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_m = 0$$

$$H_1 : \text{minimal terdapat satu } \beta_i \neq 0, i = 1, 2, \dots, m$$

Statistik uji:

$$G = -2 \ln \left[ \frac{\left(\frac{n_1}{n}\right)^{n_1} \left(\frac{n_0}{n}\right)^{n_0}}{\prod_{i=1}^n \hat{\pi}_i^{y_i} (1 - \hat{\pi}_i)^{(1-y_i)}} \right] \quad (2.13)$$

atau

$$G = 2 \left\{ \sum_{i=1}^n [y_i \ln(\hat{\pi}_i) + (1 - y_i) \ln(1 - \hat{\pi}_i)] - [n_1 \ln(n_1) + n_0 \ln(n_0) - n \ln(n)] \right\} \quad (2.14)$$

dengan

$n_0$  : banyaknya pengamatan pada kelompok  $y = 0$

$n_1$  : banyaknya pengamatan pada kelompok  $y = 1$

$n$  : jumlah total pengamatan ( $n_0 + n_1$ )

$\hat{\pi}_i = \hat{\pi}(x_i)$  : estimasi maksimum *likelihood* dari  $\pi(x_i)$

Statistik uji  $G$  mengikuti sebaran *Chi-Square* dengan taraf signifikansi  $\alpha$  dan derajat bebas  $\nu$ . Sehingga keputusan tolak  $H_0$  didapat jika nilai  $G > \chi^2_{(\alpha, \nu)}$ .

## 2.7 Pengujian Parameter Secara Parsial

Pengujian parameter secara parsial dilakukan dengan menggunakan uji Wald yang bermaksud untuk menguji signifikansi koefisien  $\beta$  secara individu (Hosmer, Lemeshow, & Studirvant, 2013). Adapun dengan hipotesis sebagai berikut:

$$H_0 : \beta_i = 0$$

$$H_1 : \beta_i \neq 0 \text{ dengan } i = 1, 2, \dots, m$$

Statistik uji :

$$W = \frac{\hat{\beta}_i}{SE(\hat{\beta}_i)} \quad (2.15)$$

dimana,

$$SE(\hat{\beta}_i) = (\text{Var}(\hat{\beta}_i))^{1/2}$$

Statistik uji  $W$  mengikuti sebaran *Chi-Square* dengan taraf signifikansi  $\alpha$  dan derajat bebas  $\nu$ . Sehingga keputusan tolak  $H_0$  didapat jika nilai  $W > \chi^2_{(\alpha, \nu)}$ .

## 2.8 Pengujian Kesesuaian Model

Pengujian kesesuaian model merupakan pengujian yang digunakan untuk memverifikasi bahwa model yang dibentuk sudah sesuai (Hosmer, Lemeshow, & Studirvant, 2013). Hipotesis yang digunakan yaitu sebagai berikut.

$$H_0 : \text{model sesuai}$$

$$H_1 : \text{model tidak sesuai}$$

Statistik yang digunakan yaitu Hosmer-Lemeshow *goodness of fit statistic* dengan persamaan sebagai berikut.

$$\hat{C} = \sum_{k=1}^g \frac{(o_{1k} - n'_k \bar{\pi}_k)'}{n'_k \bar{\pi}_k (1 - \bar{\pi}_k)} \quad (2.16)$$

dimana

$$\bar{\pi}_k = \frac{1}{n'_k} \sum_{j=1}^{c_k} m_j \hat{\pi}_j \quad (2.17)$$

$$o_{1k} = \sum_{j=1}^{c_k} y_j \quad (2.18)$$

dengan

$g$  : banyaknya kelompok

$c_k$  : jumlah pola kovariat pada kelompok ke-k

$n'_k$  : banyaknya pengamatan pada kelompok ke-k

$\bar{\pi}_k$  : estimasi rata-rata probabilitas pada kelompok ke-k

Statistik uji Hosmer-Lemeshow mengikuti sebaran *Chi-square* dengan derajat bebas  $g - 2$ . Keputusan tidak tolak  $H_0$  jika  $\hat{C} > \chi^2_{(g-2)}$ .

## 2.9 Interpretasi Koefisien Model

Perkiraan koefisien untuk variabel independen merepresentasikan *slope* (yaitu, tingkat perubahan) dari fungsi variabel dependen per unit perubahan variabel independen (Hosmer, Lemeshow, & Studirvant, 2013). *Odds ratio* dapat digunakan dalam interpretasi koefisien regresi logistik. *Odds ratio* dilambangkan dengan OR, adalah rasio peluang untuk  $x = 1$  terhadap peluang untuk  $x = 0$  ( $x$  diasumsikan sebagai  $x = 0$  dan  $x = 1$ ), dan diberikan oleh persamaan sebagai berikut.

$$\begin{aligned}
 \text{OR} &= \frac{\frac{\pi(1)}{[1-\pi(1)]}}{\frac{\pi(0)}{[1-\pi(0)]}} \\
 &= \frac{\frac{e^{\beta_0+\beta_1}}{[1-e^{\beta_0+\beta_1}]}}{\frac{e^{\beta_0}}{[1-e^{\beta_0}]}} \\
 &= \frac{e^{\beta_0+\beta_1}}{e^{\beta_0}} \\
 &= e^{(\beta_0+\beta_1)-\beta_0} \\
 &= e^{\beta_1}
 \end{aligned} \tag{2.19}$$

## 2.10 Analisis Diskriminan

Analisis diskriminan adalah teknik statistika multivariat untuk memperkirakan hubungan antar variabel dependen nonmetrik (kategori) dan serangkaian variabel

independen metrik. Analisis diskriminan mampu menangani dua kelompok atau beberapa (tiga atau lebih) kelompok. Ketika dua klasifikasi terlibat, teknik ini disebut sebagai analisis diskriminan dua kelompok. Ketika tiga atau lebih klasifikasi diidentifikasi, teknik ini disebut sebagai *Multiple-class Discriminant Analysis* (MDA) (Hair, *et al.*, 2019). Tujuan dari analisis diskriminan yaitu untuk membangun satu set diskriminan yang dapat digunakan untuk menggambarkan atau mengkarakterisasi pemisahan kelompok berdasarkan sekumpulan variabel yang direduksi, menganalisis kontribusi variabel asli terhadap pemisahan, dan mengevaluasi tingkat pemisahan (Timm, 2002). Analisis diskriminan juga memberikan dasar untuk mengklasifikasikan tidak hanya sampel yang digunakan untuk memperkirakan diskriminan fungsi tetapi juga pengamatan lain yang dapat memiliki nilai untuk semua variabel bebas. Dengan cara ini, analisis diskriminan dapat digunakan untuk mengklasifikasikan pengamatan lain ke dalam kelompok yang ditentukan (Hair, *et al.*, 2019).

### 2.10.1 Fungsi Diskriminan Dua Kelompok

Fungsi diskriminan adalah kombinasi linear dari variabel-variabel yang paling baik memisahkan kelompok. Fungsi linier dari variabel (diskriminan fungsi) digunakan untuk menggambarkan atau menjelaskan perbedaan antara dua atau lebih kelompok (Rencher, 2002). Dengan sampel  $\mathbf{x}_{11}, \mathbf{x}_{12}, \dots, \mathbf{x}_{1n_1}$  dan  $\mathbf{x}_{21}, \mathbf{x}_{22}, \dots, \mathbf{x}_{2n_2}$  dari dua kelompok dan vektor  $\mathbf{x}$  terdiri dari pengukuran pada  $m$  variabel. Fungsi diskriminan linier diberikan pada persamaan berikut (Johnson & Wichern, 2002).

$$y = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}_{pooled}^{-1} \mathbf{x} = \mathbf{a}' \mathbf{x} \quad (2.20)$$

dimana:

$$\begin{aligned} \bar{\mathbf{x}}_{1(m \times 1)} &= \frac{1}{n_1} \sum_{i=1}^{n_1} \mathbf{x}_{1i}, & \mathbf{S}_{1(m \times m)} &= \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (\bar{\mathbf{x}}_{1i} - \bar{\mathbf{x}}_1) (\bar{\mathbf{x}}_{1i} - \bar{\mathbf{x}}_1)' \\ \bar{\mathbf{x}}_{2(m \times 1)} &= \frac{1}{n_2} \sum_{i=1}^{n_2} \mathbf{x}_{2i}, & \mathbf{S}_{2(m \times m)} &= \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (\bar{\mathbf{x}}_{2i} - \bar{\mathbf{x}}_2) (\bar{\mathbf{x}}_{2i} - \bar{\mathbf{x}}_2)' \end{aligned}$$

Karena diasumsikan bahwa populasi memiliki matriks kovarian  $\Sigma$  yang sama, maka matriks kovarian sampel  $S_1$  dan  $S_2$  digabungkan untuk memperoleh estimasi  $\Sigma$  tunggal yang tak bias sebagai berikut.

$$S_{pooled} = \left[ \frac{n_1 - 1}{(n_1 - 1) + (n_2 - 1)} \right] S_1 + \left[ \frac{n_2 - 1}{(n_1 - 1) + (n_2 - 1)} \right] S_2 \quad (2.21)$$

Nilai fungsi diskriminan adalah dasar untuk menentukan suatu pengamatan dapat masuk pada kelompok yang mana dengan membandingkannya dengan nilai kritis dari fungsi diskriminan tersebut. Nilai kritis dari fungsi diskriminan dapat dihitung dengan persamaan sebagai berikut (Johnson & Wichern, 2002).

$$\begin{aligned} y_{kritis} &= \frac{\bar{y}_1 + \bar{y}_2}{2} = \frac{(\bar{x}_1 - \bar{x}_2)' S_{pooled}^{-1} \bar{x}_1 + (\bar{x}_1 - \bar{x}_2)' S_{pooled}^{-1} \bar{x}_2}{2} \\ &= \frac{1}{2} (\bar{x}_1 - \bar{x}_2)' S_{pooled}^{-1} (\bar{x}_1 - \bar{x}_2) \end{aligned} \quad (2.22)$$

Fungsi diskriminan menetapkan aturan untuk mengalokasikan sampel  $\mathbf{x}$  ke suatu kelompok, yaitu sebagai berikut.

Sampel  $\mathbf{x}_0$  dialokasikan ke dalam kelompok ke-1 jika:

$$\hat{y}_0 = (\bar{x}_1 - \bar{x}_2)' S_{pooled}^{-1} \mathbf{x}_0 \geq y_{kritis} = \frac{1}{2} (\bar{x}_1 - \bar{x}_2)' S_{pooled}^{-1} (\bar{x}_1 - \bar{x}_2)$$

dan sampel  $\mathbf{x}_0$  dialokasikan ke dalam kelompok ke-2 jika:

$$\hat{y}_0 = (\bar{x}_1 - \bar{x}_2)' S_{pooled}^{-1} \mathbf{x}_0 < y_{kritis} = \frac{1}{2} (\bar{x}_1 - \bar{x}_2)' S_{pooled}^{-1} (\bar{x}_1 - \bar{x}_2)$$

### 2.10.2 Pengujian Perbedaan Rata-rata Kelompok

Suatu fungsi diskriminan layak dibentuk apabila terdapat perbedaan nilai rata-rata diantara kelompok-kelompok yang ada. Untuk menguji perbedaan rata-rata kelompok dilakukan uji signifikansi dengan perhitungan uji  $T^2$  (Rencher, 2002). Adapun hipotesisnya sebagai berikut.

$H_0 : \mu_1 = \mu_2$  (rata-rata kelompok relatif sama)

$H_1 : \mu_1 \neq \mu_2$  (rata-rata kelompok berbeda)

Statistik uji:

$$T^2 = \left( \frac{n_1 n_2}{n_1 + n_2} \right) (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}_{pooled}^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) \quad (2.23)$$

$T^2$  akan signifikan jika  $T^2 \geq T^2_{\alpha, m, n_1 + n_2 - 2}$ , dengan taraf signifikansi  $\alpha$  dan  $m$  banyaknya variabel independen. Statistik  $T^2$  dapat diubah menjadi statistik F menggunakan:

$$\frac{n_1 + n_2 - m - 1}{(n_1 + n_2 - 2)m} T^2 = F_{m, n_1 + n_2 - m - 1} \quad (2.24)$$

## 2.11 Metode Untuk Membentuk Fungsi Diskriminan

Metode yang dapat digunakan untuk menyusun fungsi diskriminan yaitu metode seleksi *stepwise*. Metode seleksi *stepwise* adalah kombinasi dari metode seleksi *forward* dan metode seleksi *backward*. Metode seleksi *forward* dimulai dengan satu variabel, yang secara maksimal memisahkan kelompok dengan sendirinya. Maka variabelnya dimasukkan pada setiap langkah dengan memaksimalkan statistik F parsial berdasarkan Wilks' Lamda. Metode seleksi *backward* dimulai dengan semua variabel dan kemudian pada setiap langkah, variabel yang berkontribusi paling sedikit dihapus, seperti yang ditunjukkan oleh statistik F parsial (Rencher, 2002).

Maka, metode seleksi *stepwise* dimulai dengan menambahkan variabel satu per satu, dengan setiap langkah variabel diperiksa ulang untuk melihat apakah ada variabel yang telah masuk sebelumnya menjadi berlebihan dengan adanya variabel yang baru saja ditambahkan. Prosedur berhenti ketika F parsial terbesar di antara variabel yang tersedia untuk masuk gagal melebihi nilai ambang batas yang telah ditetapkan (Rencher, 2002). Persamaan Wilks' Lamda yaitu sebagai berikut.

$$\Lambda(\mathbf{x}_k | \mathbf{x}_1, \dots, \mathbf{x}_m) = \frac{\Lambda(\mathbf{x}_1, \dots, \mathbf{x}_m, \mathbf{x}_k)}{\Lambda(\mathbf{x}_1, \dots, \mathbf{x}_m)} ; k = 1, 2, \dots, g \quad (2.25)$$

dengan,

$$\Lambda(\mathbf{x}) = \frac{|\mathbf{E}|}{|\mathbf{E} + \mathbf{H}|} \quad (2.26)$$

$$\mathbf{E} = \sum_{k=1}^g \sum_{i=1}^n (\mathbf{x}_{ki} - \bar{\mathbf{x}}_{k.})(\mathbf{x}_{ki} - \bar{\mathbf{x}}_{k.})'$$

$$\mathbf{H} = n \sum_{k=1}^g (\mathbf{x}_{k.} - \bar{\mathbf{x}}_{..})(\mathbf{x}_{k.} - \bar{\mathbf{x}}_{..})'$$

dimana,

$$\mathbf{x}_{k.} = \sum_{i=1}^{n_i} \mathbf{x}_{ki}, \quad \bar{\mathbf{x}}_{k.} = \frac{\mathbf{x}_{k.}}{n_k}, \quad \bar{\mathbf{x}}_{..} = \sum_{k=1}^g \sum_{i=1}^n \mathbf{x}_{ki}$$

keterangan:

**E** : Matriks keragaman *within group* ( $m \times m$ )

**H** : Matriks keragaman *between group* ( $m \times m$ )

$g$  : banyaknya kelompok

$m$  : banyaknya variabel independen

Statistik Wilks' Lamda dalam persamaan (2.26) memiliki transformasi F sebagai berikut.

$$F = \frac{1 - \Lambda(\mathbf{x}_k | \mathbf{x}_1, \dots, \mathbf{x}_m)}{\Lambda(\mathbf{x}_k | \mathbf{x}_1, \dots, \mathbf{x}_m)} \frac{g(n-1) - m}{g-1} \quad (2.27)$$

yang berdistribusi sebagai  $F_{(g-1), (g(n-1)-m)}$ . Persamaan (2.26) dapat disebut sebagai statistik Lamda parsial dan persamaan (2.27) disebut sebagai statistik F parsial (Rencher, 2002).

## 2.12 Pengujian Normal Multivariat

Pengujian normalitas adalah pengujian yang bertujuan untuk mengetahui apakah data yang digunakan mendekati sebaran normal atau tidak. Sebaran normal multivariat adalah pengembangan dari sebaran normal univariat dengan jumlah variabel independen  $m \geq 2$  (Johnson & Wichern, 2002). Fungsi probabilitas densiti untuk sebaran normal multivariat untuk vektor acak  $\mathbf{x} = [x_1, x_2, \dots, x_m]$  adalah sebagai berikut.

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{m/2} |\Sigma|^{m/2}} e^{-\frac{1}{2}(\mathbf{x}-\mu)' \Sigma^{-1}(\mathbf{x}-\mu)} \quad (2.28)$$

Salah satu metode yang digunakan untuk menguji normalitas multivariat yaitu uji Mardia. Adapun hipotesis yang digunakan sebagai berikut.

$H_0$  : Data berdistribusi normal multivariat

$H_1$  : Data tidak berdistribusi normal multivariat

Pengujian normalitas multivariat Mardia didasarkan dari *skewness* ( $\hat{\gamma}_{1,m}$ ) dan kurtosis ( $\hat{\gamma}_{2,m}$ ) yang dihitung menggunakan persamaan sebagai berikut (Korkmaz, Goksuluk, & Zararsiz, 2014).

dengan,

$$(\hat{\gamma}_{1,m}) = \frac{1}{n^2} \sum_{a,b=1}^n M_{ab}^3 \quad (2.29)$$

$$(\hat{\gamma}_{2,m}) = \frac{1}{n^2} \sum_{a=1}^n M_{aa}^2 \quad (2.30)$$

$M_{ab} = (\bar{\mathbf{x}}_a - \bar{\mathbf{x}})' \mathbf{S}_{pooled}^{-1} (\bar{\mathbf{x}}_b - \bar{\mathbf{x}})$  yang merupakan jarak kuadrat Mahalaobis dan  $m$  adalah jumlah variabel independen. Statistik uji *skewness* adalah  $\left(\frac{n}{6}\right) \hat{\gamma}_{1,m}$  yang didekati dengan sebaran *Chi-square* dengan derajat bebas  $m(m+1)(m+2)/6$ . Sedangkan uji kurtosis adalah  $(\hat{\gamma}_{2,m})$  yang didekati dengan sebaran normal dengan rata-rata  $(m(m+2))$  dan varian  $8m(m+2)/n$  (Korkmaz, Goksuluk, & Zararsiz, 2014). Dikatakan tidak tolak  $H_0$  atau data mengikuti sebaran normal multivariat jika *p-value* untuk *skewness* dan kurtosis lebih besar dari taraf signifikan  $\alpha$ .

### 2.13 Pengujian Kesamaan Matriks Kovarian

Menurut Timm (2002), uji Box M dapat digunakan untuk menguji kesamaan matriks kovarian. Adapun hipotesisnya sebagai berikut.

$H_0 : \Sigma_1 = \Sigma_2 = \dots = \Sigma_g = \Sigma$

$H_1$  : minimal terdapat satu  $\Sigma_k \neq \Sigma$  dengan  $k = 1, 2, \dots, g$

Statistik uji :

$$X^2 = (1 - C)W \quad (2.31)$$

dimana

$$W = v_e \log|S| - \sum_{k=1}^g v_k \log|\mathbf{S}_k|$$

$$S = \sum_{k=1}^g \frac{v_k \mathbf{S}_k}{n - g}$$

$$C = \frac{2m^2 + 3m - 1}{6(m + 1)(g - 1)} \left[ \sum_{k=1}^g \frac{1}{v_k} - \frac{1}{v_e} \right]$$

dengan

$g$  : banyaknya kelompok

$m$  : banyaknya variabel independen

$\mathbf{S}_k$  : penduga tak bias dari  $\mathbf{\Sigma}_k$

$n$  :  $\sum_{k=1}^g n_k$

$n_k$  : jumlah pengamatan kelompok ke- $k$

$v_k$  :  $n_k - 1$

$v_e$  :  $n - g$

Statistik uji  $X^2$  mengikuti sebaran *Chi-square* dengan taraf signifikansi  $\alpha$  dan derajat bebas  $db = m(m + 1)(g - 1)/2$  sehingga keputusan tolak  $H_0$  jika  $X^2 > \chi_{\alpha, db}^2$ .

## 2.14 Estimasi Ketepatan Klasifikasi

Untuk menilai kemampuan mekanisme klasifikasi mengenai prediksi keanggotaan kelompok, maka bisa memakai probabilitas kesalahan klasifikasi, yang diketahui sebagai tingkat kesalahan. Dan juga bisa menggunakan pelengkapannya, yaitu tingkat klasifikasi yang benar. Perkiraan sederhana berdasarkan tingkat kesalahan dapat diperoleh dengan mencoba mekanisme klasifikasi pada kumpulan data yang sama yang telah digunakan untuk menghitung fungsi klasifikasi. Metode ini

disebut dengan resubstitusi. Setiap pengamatan dimasukkan ke fungsi klasifikasi dan ditugaskan ke kelompok, kemudian hitung jumlah klasifikasi yang benar dan jumlah kesalahan klasifikasi. Proporsi kesalahan klasifikasi yang dihasilkan dari resubstitusi disebut *APER (Apparent Error Rate)*. Hasil klasifikasi dapat dengan mudah ditampilkan dalam tabel klasifikasi atau matriks konfusi berikut (Rencher, 2002).

Tabel 1. Matriks Konfusi Hasil Klasifikasi

Aktual		Jumlah Pengamatan	Prediksi	
			Y	
			0	1
Y	0	$n_1$	$n_{11}$	$n_{12}$
	1	$n_2$	$n_{21}$	$n_{22}$

Dengan,

$n_1$  : Jumlah pengamatan aktual dari Y(0)

$n_2$  : Jumlah pengamatan aktual dari Y(1)

$n_{11}$  : Jumlah pengamatan dari Y(0) yang tepat diklasifikasikan menjadi Y(0)

$n_{12}$  : Jumlah pengamatan dari Y(0) yang salah diklasifikasikan menjadi Y(1)

$n_{21}$  : Jumlah pengamatan dari Y(1) yang salah diklasifikasikan menjadi Y(0)

$n_{22}$  : Jumlah pengamatan dari Y(1) yang tepat diklasifikasikan menjadi Y(1)

Berdasarkan matriks konfusi tersebut, dapat dihitung nilai *APER* dengan rumus sebagai berikut:

$$APER = \frac{n_{12} + n_{21}}{n_1 + n_2} \quad (2.32)$$

dan untuk menghitung tingkat ketepatan klasifikasi atau seberapa besar pengamatan dapat diklasifikasikan dengan tepat, dapat menggunakan rumus sebagai berikut:

$$\text{Tingkat Ketepatan Klasifikasi} = \frac{n_{11} + n_{22}}{n_1 + n_2} \quad (2.33)$$

atau

$$\text{Tingkat Ketepatan Klasifikasi} = 1 - \text{APER} \quad (2.34)$$

### **III. METODOLOGI PENELITIAN**

#### **3.1 Waktu dan Tempat Penelitian**

Penelitian ini dilakukan dalam semester ganjil Tahun Ajaran 2021/2022 bertempat di Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung.

#### **3.2 Data Penelitian**

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari Badan Pusat Statistik Provinsi Jawa Barat dan Jawa Tengah mengenai Tingkat Pengangguran Terbuka pada tahun 2020 beserta faktor-faktor yang diduga memiliki pengaruh. Data tersebut berjumlah 62 data berdasarkan kabupaten/kota di provinsi Jawa Barat dan Jawa Tengah.

Data Tingkat Pengangguran Terbuka (Y) berperan sebagai variabel dependen yang diklasifikasikan menjadi data biner (0 dan 1). Tingkat pengangguran terbuka kabupaten/kota dikategorikan sebagai 0 jika tingkat pengangguran terbuka kabupaten/kota lebih kecil dari tingkat pengangguran terbuka provinsi dan tingkat pengangguran terbuka kabupaten/kota dikategorikan sebagai 1 jika tingkat pengangguran terbuka kabupaten/kota lebih besar atau sama dengan tingkat pengangguran terbuka provinsi. Tingkat pengangguran terbuka provinsi Jawa Barat sebesar 10,46 dan tingkat pengangguran terbuka provinsi Jawa Tengah sebesar 6,48. Sementara untuk variabel-variabel independen yang diduga

mempengaruhi tingkat pengangguran terbuka adalah Tingkat Partisipasi Angkatan Kerja ( $X_1$ ), Indeks Pembangunan Manusia ( $X_2$ ) dan Angka Ketergantungan ( $X_3$ ).

### 3.3 Metode Penelitian

Dalam penelitian ini, proses perhitungan dilakukan dengan bantuan perangkat lunak SAS 9.4. Adapun langkah-langkah yang dilakukan yang sesuai dengan tujuan dalam penelitian ini adalah sebagai berikut.

1. Melakukan pengolahan dan penyajian data secara deskriptif pada masing-masing variabel.
2. Melakukan pengujian asumsi yaitu asumsi normal multivariat dan asumsi kesamaan matriks kovarian.
3. Mempartisi data menjadi dua yaitu data *training* dan data validasi dengan proporsi 75%-25%.
4. Melakukan analisis regresi logistik biner.
  - 1) Membangun model awal regresi logistik biner dengan menggunakan data *training* dan data secara keseluruhan.
  - 2) Melakukan pengujian signifikansi parameter secara simultan dengan menggunakan uji *Ratio Likelihood* pada model awal data *training* dan data keseluruhan.
  - 3) Melakukan pengujian signifikansi parameter secara parsial dengan menggunakan uji Wald pada model awal data *training* dan data keseluruhan.
  - 4) Membangun model baru berdasarkan hasil uji simultan dan uji parsial pada data *training* dan data keseluruhan.
  - 5) Melakukan pengujian kesesuaian model dengan menggunakan Hosmer-Lemeshow *goodness of fit statistic* pada model baru data *training* dan data keseluruhan.
  - 6) Menentukan model akhir regresi logistik biner dari data *training* dan data keseluruhan.

- 7) Menghitung nilai APER dan tingkat akurasi berdasarkan model akhir dari data yang dipartisi (data *training* dan data validasi) dan data yang tidak dipartisi (data keseluruhan).
5. Melakukan analisis diskriminan
    - 1) Melakukan pengujian persamaan vektor rata-rata kelompok dengan menggunakan uji  $T^2$ .
    - 2) Menentukan variabel yang signifikan dengan menggunakan estimasi *stepwise* berdasarkan data *training* dan data keseluruhan.
    - 3) Membentuk fungsi diskriminan berdasarkan variabel yang signifikan berdasarkan data *training* dan data keseluruhan.
    - 4) Menghitung nilai APER dan tingkat akurasi berdasarkan fungsi diskriminan yang terbentuk dari data yang dipartisi (data *training* dan data validasi) dan data yang tidak dipartisi (data keseluruhan).
  6. Membandingkan nilai APER dan tingkat akurasi data yang dipartisi (data *training* dan data validasi) dan data yang tidak dipartisi (data keseluruhan) dari metode regresi logistik biner dengan metode diskriminan.

## V. KESIMPULAN

Berdasarkan hasil dan pembahasan, didapatkan beberapa kesimpulan yaitu sebagai berikut.

1. Hasil perbandingan metode regresi logistik biner dengan metode diskriminan berdasarkan tingkat akurasi yaitu pada data yang dipartisi dengan proporsi 75%-25%, tingkat akurasi dalam mengklasifikasikan tingkat pengangguran terbuka provinsi Jawa Barat dan Jawa Tengah tahun 2020 untuk metode regresi logistik biner sebesar 67,52% sedangkan untuk metode diskriminan sebesar 68,58% (rata-rata dari tingkat akurasi data *training* dan data validasi). Dan pada data yang tidak dipartisi atau secara keseluruhan diperoleh tingkat akurasi yang sama besar untuk metode regresi logistik biner dan metode diskriminan dalam mengklasifikasikan tingkat pengangguran terbuka provinsi Jawa Barat dan Jawa Tengah tahun 2020 yaitu sebesar 67,74%.
2. Variabel independen yang memberikan pengaruh signifikan pada data yang dipartisi untuk metode regresi logistik biner yaitu hanya variabel  $X_1$  (tingkat partisipasi angkatan kerja) sedangkan untuk metode diskriminan yaitu variabel  $X_1$  (tingkat partisipasi angkatan kerja) dan variabel  $X_2$  (indeks pembangunan manusia). Dan pada data yang tidak dipartisi atau secara keseluruhan, variabel independen yang memberikan pengaruh signifikan untuk metode regresi logistik biner dan metode diskriminan yaitu sama, yakni variabel  $X_1$  (tingkat partisipasi angkatan kerja) dan variabel  $X_2$  (indeks pembangunan manusia).
3. Penerapan kedua metode (metode regresi logistik biner dan metode diskriminan) pada data tingkat pengangguran terbuka provinsi Jawa Barat dan Jawa Tengah tahun 2020 memberikan hasil yang cukup baik dengan klasifikasi yang sebenarnya.
4. Berdasarkan tingkat akurasi dalam mengklasifikasikan tingkat pengangguran terbuka provinsi Jawa Barat dan Jawa Tengah tahun 2020 menunjukkan bahwa

perbedaan antara metode regresi logistik biner dan metode diskriminan sangat kecil, bahkan tidak terdapat perbedaan jika data tidak dipartisi atau secara keseluruhan.

## DAFTAR PUSTAKA

- Al Mazah, M. M. A. 2018. Efficiency of Multivariate Logistic Regression Analysis And Discriminant Analysis In Classification Of Rich Countries According To Human Development Index. *Journal of Statistics Applications & Probability Letters*. **5**(1): 29-41.
- Gujarati, D. 2002. *Basic Econometrics*. 4<sup>th</sup> Edition. McGraw-Hill, New York.
- Hair, J. F., Black, C. W., Babin, B. J., & Anderson, R.E. 2019. *Multivariate Data Analysis*. 8<sup>th</sup> Edition. Cengage Learning, United Kingdom.
- Hosmer, D., Lemeshwo, S., & Sturdivant, R. 2013. *Applied Logistic Regression*. 3<sup>rd</sup> Edition. John Wiley & Sons, Inc., New York.
- Johnson, R. A., & Wichern, D. W. 2002. *Applied Multivariate Statistical Analysis*. 5<sup>th</sup> Edition. Prentice-Hall, Inc., New Jersey.
- Korkmaz, S., Goksuluk, D., & Zararsiz, G. 2014. MVN: An R Package for Assessing Multivariate Normality. *The R Journal*. **6**(2):151-162.
- Ndangi, R. A., Resmawan., & Djakaria, I. 2019. Perbandingan Analisis Diskriminan dan Regresi Logistik Multinomial. *Jembura Journal of Mathematics*. **1**(1): 54-63.
- Ranchar, A. C. 2002. *Methods of Multivariate Analysis*. 2<sup>nd</sup> Edition. John Wiley & Sons, Inc., New York.
- Ryan, T.P. 1997. *Modern Regression Methods*. John Wiley & Sons, Inc., New York.

Santoso, S. 2017. *Statistik Multivariat dengan SPSS*. Elek Media Komputindo, Jakarta.

Timm, N. H. 2002. *Applied Multivariate Analysis*. Springer-Verlag New York, Inc., New York.