

Skripsi Statistika

**KLASIFIKASI TINGKAT KESEJAHTERAAN
MASYARAKAT SULAWESI SELATAN DENGAN METODE
KUARTIL MENGGUNAKAN FUNGSI ANALISIS
DISKRIMINAN LINIER**

SKRIPSI



NURUL HIJRAH

H051181503

**PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN
MAKASSAR
DESEMBER 2022**

**KLASIFIKASI TINGKAT KESEJAHTERAAN
MASYARAKAT SULAWESI SELATAN DENGAN METODE
KUARTIL MENGGUNAKAN FUNGSI ANALISIS
DISKRIMINAN LINIER**

SKRIPSI

**Diajukan sebagai salah satu syarat memperoleh gelar Sarjana Sains pada
Program Studi Statistika Departemen Statistika Fakultas Matematika dan
Ilmu Pengetahuan Alam Universitas Hasanuddin**

NURUL HIJRAH

H051181503

**PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**

UNIVERSITAS HASANUDDIN

MAKASSAR

DESEMBER 2022

LEMBAR PERNYATAAN KEOTENTIKAN

Saya yang bertanda tangan di bawah ini menyatakan dengan sungguh-sungguh bahwa skripsi yang saya buat dengan judul:

Klasifikasi Tingkat Kesejahteraan Masyarakat Sulawesi Selatan Dengan Metode Kuartil Menggunakan Analisis Diskriminan Linier

adalah benar hasil karya saya sendiri, bukan hasil plagiat dan belum pernah dipublikasikan dalam bentuk apapun

Makassar, 02 Desember 2022



Nurul Hijrah

NIM H051181505

**KLASIFIKASI TINGKAT KESEJAHTERAAN MASYARAKAT
SULAWESI SELATAN DENGAN METODE KUARTIL
MENGUNAKAN ANALISIS DIKRIMINAN LINIER**

Disetujui Oleh:

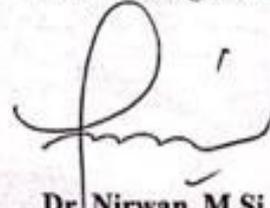
Pembimbing Utama



Drs. Raupong, M.Si.

NIP. 19621015 198810 1 001

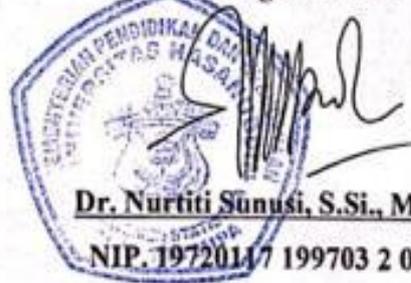
Pembimbing Pertama



Dr. Nirwan, M.Si.

NIP. 19630306 198702 1 002

Ketua Program Studi



Dr. Nurtiti Sunusi, S.Si., M.Si.

NIP. 19720117 199703 2 002

Pada 02 Desember 2022

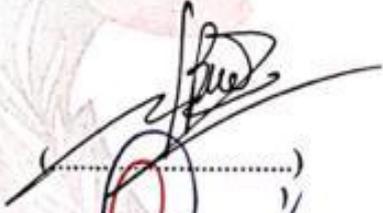
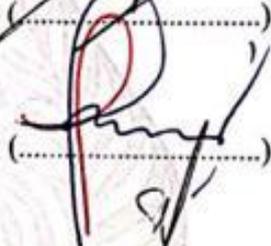
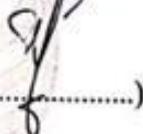
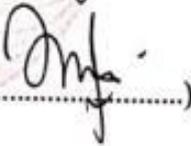
HALAMAN PENGESAHAN

Skripsi ini diajukan oleh:

Nama : Nurul Hijrah
NIM : H051181503
Program Studi : Statistika
Judul Skripsi : Klasifikasi Tingkat Kesejahteraan Masyarakat Sulawesi Selatan dengan Metode Kuartil Menggunakan Analisis Diskriminan Linier

Telah berhasil dipertahankan di hadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Sains pada Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

DEWAN PENGUJI

1. Ketua : Drs. Raupong, M.Si. 
2. Sekretaris : Dr. Nirwan, M.Si. 
3. Anggota : Dr. Dr. Georgina Maria Tinungki, M.Si. 
4. Anggota : Anisa, S.Si., M.Si. 

Ditetapkan di : Makassar

Tanggal : 02 Desember 2022

KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabarakatuh

Segala puji bagi Allah SWT, tuhan semesta alam yang senantiasa melimpahkan kasih-Nya. karena pertolongan-Nya, akhirnya skripsi dengan judul “Klasifikasi Tingkat Kesejahteraan Masyarakat dengan Metode Kuartil Menggunakan Analisis Diskriminan Linier” yang disusun sebagai salah satu syarat akademik untuk meraih gelar sarjana pada Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin ini dapat dirampungkan.

Penulisan skripsi ini tidak luput dari bantuan berbagai pihak. Oleh karena itu, penulis mengucapkan terima kasih yang sebesar-besarnya kepada orang tua penulis, ayahanda tercinta **Alm. Ali Senong S.Pd** yang selalu cinta dan dukungan kepada penulis dari kejauhan yang tak terlihat dan ibunda tersayang **Hj. Kadrang** yang senantiasa memberikan doa, bantuan, dukungan setiap langkah perjalanan penulis dalam menuntut ilmu. Terima kasih juga penulis ucapkan kepada kakak tersayang **Ade Lestari** dan **Andi Ilham** yang senantiasa membantu dan menyertakan penulis dalam setiap doanya.

Penghargaan yang tulus dan ucapan terima kasih yang sebesar-besarnya juga penulis ucapkan kepada:

1. **Bapak Prof. Dr. Ir. Jamaluddin Jompa, M.Sc** selaku Rektor Universitas Hasanuddin beserta seluruh jajarannya.
2. **Bapak Dr, Eng. Amiruddin**, selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin beserta seluruh jajarannya.
3. **Ibu Dr. Nurtiti Sunusi, M.Si**, selaku ketua Departemen Statistika, dan **Bapak Andi Kresna Jaya, S.Si., M.Si** selaku Sekretaris Departemen Statistika yang telah membantu dan membekali ilmu kepada penulis selama menjadi mahasiswa di Departemen Statistika.
4. **Bapak Drs. Raupong, M.Si**, selaku Dosen Pembimbing Utama sekaligus Penasehat Akademik dan **Bapak Dr. Nirwan, M.Si**, selaku dosen Pembimbing Pertama yang senantiasa sabar dalam menuntun penulis selama menyusun tugas akhir ini dan telah meluangkan waktu ditengah berbagai kesibukan untuk

membimbing, berbagi ilmu dan memberikan dukungan kepada penulis dalam menyelesaikan tugas akhir ini.

5. **Ibu Dr. Dr. Georgina Maria Tinungki, M.Si** dan **Ibu Anisa, S.Si., M.Si** selaku penguji. Terima kasih atas segala masukan bantuan, nasehat serta motivasi yang telah diberikan kepada penulis selama menjalani pendidikan di Departemen Statistika.
6. Terima kasih seluruh Staf Pegawai dan Dosen di Jurusan Statistika yang telah membantu proses administrasi selama penulis menyelesaikan tugas akhir. Ucapan terima kasih juga penulis sampaikan kepada orang-orang yang telah berperan besar serta istimewa kepada:
 1. Sahabat penulis **Yayung Kangkang** yang selalu menjadi tempat terbaik dalam mencurahkan isi hati baik keadaan senang maupun sedih, tetap menjadi *support system* terbaik *love u so muchh*.
 2. Sahabat penulis **Juana** terima kasih selalu menemani dalam keadaan apapun *hop u always there*. Sahabat penulis **Sapite** terima kasih karena selalu membuat *mood* penulis *always happy*.
 3. Sahabat penulis **Emi, Hapise,** dan **mampu** terima kasih telah menjadi pendengar yang baik, selalu mendukung satu sama lain, dan telah berjuang bersama-sama selama masa perkuliahan.
 4. Teman-teman Netijen **Kak Pan, Dian, Ika, Ren,** dan **Poet** terima kasih atas bantuan, dukungan dan kesenangan yang diberikan kepada penulis selama menghadapi masa-masa terindah dan tersulit diperkuliahan, terima kasih juga karena telah bertahan dalam 4 tahun terakhir ini dan harus untuk selamanya *I Love You Gess*. Teman-teman Koperasi Bersatu **Haksare, Margarita, Mentu, Restu, Kejia, Acca, Aksa, Rael** yang telah memberikan semangat dan bantuannya kepada penulis selama menyelesaikan tugas akhir ini.
 5. Teman-teman Lesgobel **Gusti, Ramma, Aldi** terima kasih atas bantuan pembuatan PPTnya yang sangat membantu penulis dalam menyelesaikan tugas akhir.
 6. Teman-temaan Bandit **Annise** dan **Laija** yang selalu bisa diajak untuk menghibur penulis saat menyusun tugas akhir ini.

7. Gadis-gadis Posko 5 **Ratna, Jukulup, Husna** yang selalu menjadi tempat melimpahkan kedumbaan *keep solid and always* kebersamai. Gembul **Hisyam, Nugi, Parrul, Toke, Pebi, Kak Jul** terima kasih selalu mendukung dan membantu penulis dalam menyelesaikan tugas akhir.
8. Keluarga Besar **Statistika 2018** yang tak sempat disebut namanya, terima kasih telah menjadi keluarga yang senantiasa menemani dalam seluruh kegiatan akademik maupun non-akademik.
9. Last but not least terima kasih kepada diri sendiri karena telah berjuang sampai titik ini, terima kasih telah mewujudkan salah satu impian terakhir papamu dan terima kasih karena selalu bangkit dengan senyuman maupun air mata yang kebersamai, *greatfull and so proud of my self*.

Penulis menyadari bahwa masih banyak kekurangan dalam skripsi ini, untuk itu penulis memohon maaf dengan segala kerendahan hati. Akhir kata semoga tulisan ini dapat memberikan manfaat untuk pembaca.

Makassar, 02 Desember 2022

Nurul Hijrah

**PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK
KEPENTINGAN AKADEMIK**

Sebagai civitas akademik Universitas Hasanuddin, saya yang bertanda tangan di bawah ini:

Nama : Nurul Hijrah
NIM : H051181503
Program Studi : Statistika
Departemen : Statistika
Fakultas : Matematika dan Ilmu Pengetahuan Alam
Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Hasanuddin **Hak Bebas Royalti Non-eksklusif** (*Non-exclusive Royalty- Free Right*) atas tugas akhir saya yang berjudul:

“Klasifikasi Tingkat Kesejahteraan Masyarakat Sulawesi Selatan Dengan Metode Kuartil Menggunakan Analisis Diskriminan Linier”

Beserta perangkat yang ada (jika diperlukan). Terkait dengan hal di atas, maka pihak universitas berhak menyimpan, mengalih-media/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan memublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di Makassar pada tanggal, 02 Desember 2022

Yang menyatakan

(Nurul Hijrah)

ABSTRAK

Keadaan Indonesia semakin memprihatinkan khususnya untuk perkembangan ekonomi. Hal tersebut dapat kita lihat dengan masih banyaknya masyarakat dengan ekonomi rendah di sejumlah daerah. Untuk itu diperlukan pemerataan pendapatan di setiap daerah agar kesejahteraan dapat dirasakan oleh setiap masyarakat. Berdasarkan hal tersebut, pemerintah sebaiknya melakukan penetapan pendapatan tingkat kesejahteraan masyarakat di setiap daerah. Penelitian ini bertujuan untuk mengklasifikasikan tingkat kesejahteraan masyarakat Sulawesi Selatan tahun 2020 serta mengklasifikasikan hasil pengamatan baru tingkat kesejahteraan masyarakat Sulawesi Selatan tahun 2021. Pengklasifikasian ini dilakukan dengan metode perhitungan kuartil menggunakan analisis diskriminan linier. Fungsi diskriminan linier yang diperoleh adalah $Y_1 = -0,5993X_1 + 0,2515X_2 - 0,7599X_3 - 0,0008X_4$ dengan ketepatan hasil klasifikasi tingkat kesejahteraan masyarakat Sulawesi Selatan sebesar 82,79%. Fungsi analisis diskriminan linier juga mengakibatkan perubahan klasifikasi tingkat kesejahteraan masyarakat kabupaten/kota Sulawesi Selatan tahun 2020 jika dibandingkan dengan klasifikasi awal dengan menggunakan perhitungan kuartil. Hasil yang diperoleh adalah fungsi diskriminan linier yang dapat mengklasifikasikan tingkat kesejahteraan masyarakat di Sulawesi Selatan pada tahun 2020 dan tahun 2021.

Kata Kunci : Analisis Diskriminan Linier, Klasifikasi, Kuartil, Tingkat Kesejahteraan Masyarakat

ABSTRACT

The situation in Indonesia is increasingly concerning, especially for economic development. We can see this with the number of community with low economies in a number of regions. For this reason, it is necessary to equalize income in each region so that welfare can be felt by every community. Based on this research, the government should determine the income level of welfare of the people in each region. This study aims to classify the welfare level of the community of South Sulawesi in 2020 and classify the results of new observations on the welfare level of the community of South Sulawesi in 2021. This classification is carried out by the quartile calculation method using linear discriminant analysis. The linear discriminant function obtained is $Y_1 = -0,5993X_1 + 0,2515X_2 - 0,7599X_3 - 0,0008X_4$ with the accuracy of the classification results for the welfare level of the people of South Sulawesi of 82,79%. The linear discriminant analysis function also resulted in a change in the classification of the welfare level of the South Sulawesi district/city community in 2020 when compared to the initial classification using quartile calculations. The results obtained are linear discriminant function that can be classify the level of community welfare in South Sulawesi in 2020 and 2021.

Keywords : Classification, community welfare level, linear discriminant analysis, quartile

DAFTAR ISI

HALAMAN SAMPUL	i
HALAMAN JUDUL.....	ii
LEMBAR PERNYATAAN KEOTENTIKAN	iii
HALAMAN PERSETUJUAN PEMBIMBING.....	iv
HALAMAN PENGESAHAN	v
KATA PENGANTAR	vi
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI.....	ix
ABSTRAK.....	x
ABSTRACT	xi
DAFTAR ISI	xii
DAFTAR TABEL	xiv
DAFTAR GAMBAR	xv
DAFTAR LAMPIRAN	xvi
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
BAB II TINJAUAN PUSTAKA	4
2.1 Kesejahteraan Masyarakat	4
2.1.1 Indeks Pembangunan Masyarakat	4
2.1.2 Angka Harapan Hidup.....	5
2.1.3 Harapan Lama Sekolah.....	5
2.1.4 Rata-rata Lama Sekolah	6
2.2 Perhitungan Kuartil.....	7
2.3 Uji Asumsi	12
2.3.1 Distribusi Normal Multivariat	13
2.3.2 Uji Vektor Rata-rata	14
2.3.3 Uji Kesamaan Matriks Variansi-kovariansi	15

2.4	Pengklasifikasian dengan Menggunakan Analisis Diskriminan Linier	8
2.5	Uji Signifikansi Fungsi Linier	16
2.6	Uji Keakuratan Prediksi Klasifikasi	16
BAB III METODOLOGI PENELITIAN.....		18
3.1	Sumber Data	18
3.2	Variabel Penelitian.....	18
3.3	Langkah-langkah Analisis	19
BAB IV HASIL DAN PEMBAHASAN		21
4.1	Pengklasifikasian Data Berdasarkan Perhitungan Kuartil	21
4.2	Statistik Deskriptif	21
4.3	Uji Asumsi	22
4.3.1	Uji Distribusi Normal Multivariat	22
4.3.2	Uji Vektor Rata-rata	23
4.3.3	Uji Kesamaan Matriks Variansi-Kovariansi	24
4.4	Analisis Diskriminan Linier.....	26
4.5	Uji Keakuratan Prediksi Klasifikasi	28
4.6	Pengklasifikasian Data Baru	29
4.7	Pengklasifikasian Data Berdasarkan Perhitungan BPS	32
BAB V PENUTUP		35
5.1	Kesimpulan.....	35
5.2	Saran.....	35
DAFTAR PUSTAKA		36
LAMPIRAN		38

DAFTAR TABEL

Tabel 2.1 Struktur Data dalam Analisis Diskriminan Linier	8
Tabel 3.1 Variabel Penelitian.....	18
Tabel 4.1 Statistik Deskriptif Tingkat Kesejahteraan Masyarakat Kabupaten/kota Sulawesi Selatan Tahun 2020.....	22
Tabel 4.2 Hasil Klasifikasi Tingkat Kesejahteraan Masyarakat Kabupaten/Kota Sulawesi Selatan Tahun 2020 dengan Menggunakan Analisis Diskriminan Linier	27
Tabel 4.3 Perbandingan Pengklasifikasian Data Baru Tingkat Kesejahteraan Masyarakat Kabupaten/kota Sulawesi Selatan Tahun 2021 dan Tahun 2020	31
Tabel 4.4 Hasil Klasifikasi Kesejahteraan Masyarakat Kabupaten/Kota Sulawesi Selatan Tahun 2020 Berdasarkan Pengelompokan BPS	33

DAFTAR GAMBAR

Gambar 4. 1 QQ-Plot Tingkat Kesejahteraan Masyarakat Sulawesi Selatan Tahun
2020 23

DAFTAR LAMPIRAN

Lampiran 1. Data Penelitian	39
Lampiran 2. Pengelompokan Data IPM Tahun 2020 Berdasarkan Perhitungan Kuartil	40
Lampiran 3. Perhitungan Jarak Mahalanobis	41
Lampiran 4. Klasifikasi Data Baru Tingkat Kesejahteraan Masyarakat Kabupaten/kota Sulawesi Selatan Tahun 2021	42
Lampiran 5. Syntax dan Output Program SAS Studio	43

BAB I

PENDAHULUAN

1.1 Latar Belakang

Indonesia merupakan salah satu negara yang saat ini masih tergolong negara berkembang. Hal ini menjadikan kesejahteraan ekonomi masih sangat dibutuhkan oleh masyarakat Indonesia, untuk itu diperlukan pemerataan pendapatan di setiap daerah agar kesejahteraan dapat dirasakan oleh setiap masyarakat. Berdasarkan hal tersebut, pemerintah sebaiknya melakukan penetapan pendapatan tingkat kesejahteraan masyarakat di setiap daerah.

Seperti yang dijelaskan dalam UU No.11 Tahun 2009 kesejahteraan masyarakat adalah suatu kondisi terpenuhinya kebutuhan material, spiritual, dan sosial warga negara agar dapat hidup layak dan mampu mengembangkan diri, sehingga dapat melaksanakan fungsi sosialnya. Pada tahun 2020 kondisi kesejahteraan masyarakat di Sulawesi Selatan semakin membaik meski berada pada masa pandemi covid-19. Salah satu parameter yang dapat digunakan untuk menentukan status tingkat kesejahteraan masyarakat di suatu wilayah ialah Indeks Pembangunan Manusia (BPS, 2011).

Menurut Badan Pusat Statistik (BPS), IPM dikelompokkan menjadi 4 kategori yaitu sangat tinggi, tinggi, sedang, dan rendah. Untuk data IPM Sulawesi Selatan tahun 2020 berdasarkan pengelompokan BPS hanya memenuhi dua kategori yaitu tinggi dan sedang. IPM dipengaruhi oleh beberapa faktor diantaranya Angka harapan hidup (AHH) adalah jumlah rata-rata usia yang diperkirakan pada seseorang atas dasar angka kematian yang cenderung tidak berubah dimasa mendatang. Faktor yang kedua yaitu harapan lama sekolah (HLS) yang didefinisikan sebagai lamanya sekolah (dalam tahun) yang diharapkan akan dirasakan oleh anak pada umur tertentu di masa mendatang. Faktor yang ketiga adalah rata-rata lama sekolah (RLS) juga didefinisikan sebagai rata-rata jumlah tahun yang ditempuh oleh penduduk untuk usia 25 tahun ke atas dengan asumsi pada umur 25 tahun proses pendidikan sudah berakhir. Faktor yang terakhir adalah pengeluaran per kapita (PPP) merupakan jumlah dikeluarkan secara tunai oleh pemerintah daerah baik secara rutin yang kemudian dinamakan sebagai belanja rutin maupun belanja modal (BPS,2021).

Tingkat kesejahteraan masyarakat dapat digolongkan berdasarkan nilai indeks pembangunan manusia dengan perhitungan kuartil. Kuartil merupakan ukuran letak yang membagi data yang telah diukur dan diurutkan dari $x_1 \leq x_2 \leq \dots \leq x_n$ yang membagi data menjadi empat bagian yang sama besar. Sama halnya dengan ukuran-ukuran yang lain, kuartil digunakan untuk mempermudah pengolahan data meliputi perhitungan dan pengelompokan (Sugiono, 2014).

Analisis diskriminan linier adalah salah satu metode dalam statistika yang dapat digunakan untuk mengevaluasi klasifikasi objek. Sehingga metode analisis diskriminan linier menjadi salah satu metode yang dapat digunakan dalam menganalisis status tingkat kesejahteraan masyarakat berdasarkan variabel tersebut. Analisis diskriminan linier merupakan salah satu dari analisis multivariat dengan metode dependensi. Tujuan dari analisis diskriminan linier adalah untuk mendapatkan suatu fungsi yang disebut dengan fungsi diskriminan, dimana fungsi ini dapat memisahkan objek sesuai dengan kelompoknya. Fungsi diskriminan ini juga dapat digunakan untuk memprediksi kelompoknya dari objek baru yang diamati (Sharma, 1996).

Analisis diskriminan digunakan dalam beberapa penelitian terkait klasifikasi seperti yang dilakukan oleh Hudri dan Harison (2021) Analisis Diskriminan Linier Untuk Klasifikasi Kabupaten/Kota Di Provinsi Riau Berdasarkan Indikator Indeks Pembangunan Manusia dengan hasil penelitian yaitu fungsi diskriminan linear yang terbentuk signifikan dan akurat dengan tingkat validasi sebesar 91.6% dalam mengklasifikasikan kabupaten/kota di provinsi Riau. Selain itu, Kartikawati, Mukid, dan Ispriyanti (2013) Perbandingan Analisis Diskriminan Linier Klasik Dan Analisis Diskriminan Linear Robust Untuk Pengklasifikasian Kesejahteraan Masyarakat Kabupaten/Kota Di Jawa Tengah berdasarkan indikator kesejahteraan masyarakat dengan hasil penelitian yaitu ketepatan pengklasifikasian analisis diskriminan linear klasik dan robust yaitu sebesar 77,14% dalam mengklasifikasikan kabupaten/kota di provinsi Jawa Tengah.

Berdasarkan uraian tersebut, penulis akan membahas mengenai analisis diskriminan yang akan dituangkan dalam penelitian yang berjudul **“Klasifikasi Tingkat Kesejahteraan Masyarakat Sulawesi Selatan dengan Metode Kuartil Menggunakan Analisis Diskriminan Linier”**.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah dalam penelitian ini adalah:

1. Bagaimana fungsi analisis diskriminan linier dalam mengklasifikasikan atau mengelompokkan tingkat kesejahteraan masyarakat Sulawesi Selatan tahun 2020 berdasarkan perhitungan kuartil?
2. Bagaimana pengklasifikasian pengamatan baru tingkat kesejahteraan masyarakat Sulawesi Selatan tahun 2021?

1.3 Batasan Masalah

Adapun batasan masalah pada penelitian ini adalah sebagai berikut:

1. Metode yang digunakan pada penelitian ini adalah analisis diskriminan linier dan pengklasifikasian berdasarkan perhitungan kuartil.
2. Data yang digunakan adalah data tingkat kesejahteraan masyarakat kabupaten/kota Sulawesi Selatan pada tahun 2020.

1.4 Tujuan Penelitian

Maksud dan tujuan penyusunan proposal ini adalah sebagai berikut:

1. Memperoleh fungsi diskriminan linier dalam mengklasifikasikan tingkat kesejahteraan masyarakat Sulawesi Selatan.
2. Memperoleh pengklasifikasian pengamatan baru tingkat kesejahteraan masyarakat Sulawesi Selatan tahun 2021.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat yaitu:

1. Dapat menambah pengetahuan dan wawasan terkait penerapan metode Analisis Diskriminan.
2. Penelitian ini diharapkan dapat menjadi acuan untuk peneliti yang lain untuk mengkaji lebih dalam lagi.

BAB II TINJAUAN PUSTAKA

2.1 Kesejahteraan Masyarakat

Kesejahteraan merupakan titik ukur bagi suatu masyarakat bahwa telah berada pada kondisi sejahtera (Mulia & Saputra 2020). Kesejahteraan tersebut dapat diukur dari kesehatan, keadaan ekonomi, kebahagiaan dan kualitas hidup rakyat, Pandangan masyarakat umum, dalam keluarga yang sejahtera maka mampu menyekolahkan anggota keluarganya hingga setinggi mungkin. Sama halnya jika semakin tinggi tingkat pendidikan seseorang maka akan membawa keluarganya semakin sejahtera karena mendapatkan timbal balik seperti pekerjaan yang mapan dan pendapatan yang mencukupi (Mulia & Saputra, 2020). Kondisi sejahtera yang dimaksud juga menunjuk pada kesejahteraan sosial, yaitu tercukupinya kebutuhan material dan non material (Suud, 2006).

Bagi masyarakat, kondisi sejahtera diartikan sebagai kondisi dimana seseorang dapat memenuhi kebutuhan pokok, baik itu kebutuhan akan makanan, pakaian, tempat tinggal, air minum yang bersih serta kesempatan untuk melanjutkan pendidikan dan memiliki pekerjaan yang memadai yang dapat menunjang kualitas hidupnya sehingga hidupnya bebas dari kemiskinan, kebodohan, ketakutan, atau kekhawatiran (Saifudin, 2019).

2.1.1 Indeks Pembangunan Masyarakat

Pada Tahun 1990, *United Nations Development Programme* (UNDP) telah menerbitkan suatu indikator yaitu Indeks Pembangunan Manusia (IPM) untuk mengukur kesuksesan pembangunan dan kesejahteraan suatu negara. Indeks Pembangunan Manusia (IPM) dilihat berdasarkan tiga dimensi yaitu: angka harapan hidup, angka harapan lama sekolah dan rata-rata lama sekolah mengukur pendidikan dan terakhir daya beli mengukur standar hidup. Ketiga indikator tersebut saling mempengaruhi satu sama lain. Maka semakin tinggi indeks pembangunan manusia di suatu daerah menandakan bahwa kesejahteraan masyarakat daerah tersebut meningkat. (UNDP, 1990).

Indeks Pembangunan Manusia juga menjelaskan tentang cara manusia memilih peluang untuk mengakses hasil dari proses pembangunan suatu wilayah, sebagai

tanda bagian dari hak yang harus di terima seperti memperoleh pendapatan sesuai standar, pendidikan yang layak, fasilitas kesehatan yang memadai serta pemerataan kesejahteraan. Berdasarkan kajian mengenai indeks pembangunan manusia, menurut UNDP IPM untuk mengetahui capaian indeks pembangunan manusia antar wilayah dapat dilihat melalui pengelompokkan ke dalam beberapa kategori, yaitu (IPM Sulsel, 2021):

Sangat tinggi : $IPM \geq 80$

Tinggi : $70 \leq IPM < 80$

Sedang : $60 \leq IPM < 70$

Rendah : $IPM < 60$

2.1.2 Angka Harapan Hidup

Angka Harapan Hidup (AHH) didefinisikan sebagai rata-rata perkiraan banyak tahun yang dapat ditempuh oleh seseorang sejak lahir. Umur harapan hidup dijadikan indikator dalam mengukur kesehatan suatu individu di suatu daerah. Umur harapan hidup dihitung menggunakan pendekatan tak langsung (*indirect estimation*). Ada dua jenis data yang digunakan dalam penghitungan umur harapan hidup yaitu anak lahir hidup dan anak masih hidup. Selain itu usia harapan hidup juga dapat panjang jika status kesehatan, gizi, dan lingkungan yang baik (IPM Sulsel, 2021).

2.1.3 Harapan Lama Sekolah

Badan Pusat Statistik 2021 menyatakan angka harapan lama sekolah (HLS) adalah sebagai lamanya sekolah (dalam tahun) yang diharapkan akan dirasakan oleh anak pada umur tertentu dimasa mendatang. Pentingnya angka harapan lama sekolah untuk mengetahui kondisi pembangunan sistem pendidikan diberbagai jenjang yang ditujukan dalam bentuk lamanya pendidikan yang diharapkan dapat dicapai oleh setiap anak (IPM Sulsel, 2021). Angka harapan lama sekolah dihitung untuk penduduk berusia 7 tahun keatas. Adapun rumus perhitungan angka harapan lama sekolah berdasarkan sebagai berikut:

$$HLS_a^t = FK \times \sum_{i=a}^n \frac{E_i^t}{P_t^i}$$

dengan :

HLS_a^t : Harapan lama sekolah pada umur a di tahun t

- FK : Faktor koreksi
 E_i^t : Jumlah penduduk yang berusia i yang bersekolah pada tahun t
 i : Usia ($a, a + 1, \dots n$)
 P_t^i : Penduduk usia i

2.1.4 Rata-rata Lama Sekolah

Rata-rata Lama Sekolah (RLS) atau *Mean Years of Schooling* (MYS) didefinisikan sebagai jumlah tahun yang digunakan oleh penduduk dalam menjalani pendidikan formal. Rata-rata lama sekolah mengindikasikan makin tingginya pendidikan yang dicapai oleh masyarakat di suatu daerah. Semakin tinggi rata-rata lama sekolah berarti semakin tinggi jenjang pendidikan yang dijalani. Asumsi yang berlaku secara umum bahwa semakin tinggi tingkat pendidikan seseorang maka semakin tinggi pula kualitas seseorang, baik pola pikir maupun polanya. Tobing mengemukakan bahwa orang yang memiliki tingkat pendidikan lebih tinggi, diukur dengan lamanya waktu sekolah akan memiliki pekerjaan dan upah yang lebih baik dibanding dengan orang yang pendidikannya lebih rendah (Siskawati dkk, 2021).

RLS juga digunakan pada IPM metode lama dan metode baru untuk mengukur dimensi pendidikan. Tetapi terdapat perbedaan mendasar dalam definisi RLS ini. Pada IPM metode lama, RLS dihitung untuk penduduk usia 15 tahun ke atas. Sedangkan pada IPM metode baru, cakupan penduduk yang dihitung RLS adalah penduduk berusia 25 tahun ke atas. RLS yang dihitung untuk usia 25 tahun ke atas dengan asumsi pada umur 25 tahun proses pendidikan sudah berakhir. Selain itu, penghitungan RLS pada usia 25 tahun ke atas juga mengikuti standar internasional yang digunakan oleh UNDP (IPM Sulsel, 2021).

$$RLS = \frac{1}{N} \times \sum_{i=1}^N x_i$$

dengan :

- RLS : Rata-rata lama sekolah penduduk usia 25 tahun ke atas
 x_i : Lama sekolah penduduk ke i yang berusia
 N : Jumlah penduduk usia 25 tahun ke atas

2.1.5 Pengeluaran Per Kapita

Pengeluaran per kapita merupakan jumlah yang dikeluarkan secara tunai oleh pemerintah daerah baik secara rutin yang kemudian dinamakan sebagai belanja rutin maupun belanja modal. Selain itu, jumlah penduduk yang tinggi dan menjadi beban yang tinggi bagi pemerintah daerah dapat berpengaruh terhadap kinerja keuangan pemerintah daerah. Halim dan Damayanti menyatakan bahwa jumlah belanja modal yang tinggi mengindikasikan bahwa pemerintah daerah lebih banyak melakukan pengeluaran daerah akan dirasakan pada beberapa tahun setelah terjadi belanja modal tersebut. Pengaruh pengeluaran ini telah dibuktikan oleh Steven dan McGowen pada tahun 1983 terhadap kinerja keuangan pemerintah. Jadi, Pengeluaran per kapita merupakan proporsi antara jumlah total pengeluaran pemerintah daerah dengan jumlah penduduk di suatu pemerintah daerah tersebut. Adapun rumus perhitungan pengeluaran per kapita sebagai berikut (IPM, 2021):

$$PPP = \frac{\text{Jumlah Pengeluaran Daerah}}{\text{Jumlah Penduduk}}$$

2.2 Perhitungan Kuartil

Kuartil adalah ukuran letak yang membagi data yang telah diukur dan diurutkan menjadi empat bagian yang sama besar (Pratikno, Prastiwi, dan Rahmawati, 2020). Kuartil digunakan untuk mempermudah pengolahan data meliputi perhitungan dan pengelompokan. Kuartil pertama (Q_1) memiliki kuantitas 25%, kuartil ketiga (Q_3) memiliki kuantitas 75%, dan kuartil kedua (Q_2) atau kuartil tengah memiliki kuantitas 50% tidak lain adalah median dimana membagi data menjadi dua bagian yang sama besar. Dalam menentukan letak kuartil biasanya digunakan rumus sebagai berikut (Sugiono, 2014):

$$\text{Letak } Q_i \text{ pada data ke : } \frac{i(n+1)}{4} \quad (2.1)$$

$$\text{Nilai Letak } Q_i = \frac{3}{4} \times \text{data ke } (\text{int}(\text{Letak } Q_i) + \frac{1}{4} \times (\text{data ke } (1 + \text{int}(\text{Letak } Q_i))) \quad (2.2)$$

dengan

$i = 1, 2, \text{ dan } 3$

n : Banyaknya data

2.3 Pengklasifikasian dengan Menggunakan Analisis Diskriminan Linier

Analisis diskriminan merupakan suatu metode statistika yang digunakan untuk mengklasifikasikan suatu individu atau objek ke dalam suatu kelompok yang telah ditentukan sebelumnya berdasarkan variabel-variabel bebasnya (Dillon dan Goldstein, 1984). Tujuan dari analisis diskriminan yaitu untuk mengidentifikasi variabel-variabel yang mampu membedakan antara kedua kelompok (Ghozali I, 2016). Dalam analisis analisis diskriminan terdapat beberapa asumsi yang harus dipenuhi yaitu uji distribusi normal multivariat, uji vektor rata-rata, dan uji kesamaan matriks variansi-kovariansi. Hasil dari analisis diskriminana ini yaitu fungsi yang dapat digunakan untuk tujuan klasifikasi.

Klasifikasi berasal dari bahasa latin yaitu *classis* yang artinya pembagian sesuatu menurut kelas-kelas. Istilah klasifikasi menunjuk kepada sebuah metode untuk menyusun data secara sistematis (Sulistyo Basuki, 1997). Dalam dunia statistika pengklasifikasian banyak digunakan dalam berbagai metode salah satunya analisis diskriminan.

Struktur data yang digunakan dalam analisis diskriminan linier adalah sebagai berikut:

Tabel 2.1 Struktur Data dalam Analisis Diskriminan Linier

Kelompok	Variabel					
	X_1	X_2	X_3	...	X_p	
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	x_{111}	x_{121}	x_{131}	...	x_{1p1}	
	x_{211}	x_{221}	x_{231}	...	x_{2p1}	
	x_{311}	x_{321}	x_{331}	...	x_{3p1}	
	⋮	⋮	⋮	⋮	⋮	
	x_{n_111}	x_{n_121}	x_{n_131}	...	x_{n_1p1}	
Rataan	$\bar{x}_{.11}$	$\bar{x}_{.21}$	$\bar{x}_{.31}$...	$\bar{x}_{.p1}$	$\bar{x}_{..1}$
2	x_{112}	x_{122}	x_{132}	...	x_{1p2}	
	x_{212}	x_{222}	x_{232}	...	x_{2p2}	
	x_{312}	x_{322}	x_{332}	...	x_{3p2}	

(1)	(2)	(3)	(4)	(5)	(6)	(7)
	⋮	⋮	⋮	⋮	⋮	
	$x_{n_2 12}$	$x_{n_2 22}$	$x_{n_2 32}$	⋯	$x_{n_2 p2}$	
Rataan	$\bar{x}_{.12}$	$\bar{x}_{.22}$	$\bar{x}_{.32}$	⋯	$\bar{x}_{.p2}$	$\bar{x}_{..2}$
;3	x_{113}	x_{123}	x_{133}	⋯	x_{1p3}	
	x_{213}	x_{223}	x_{233}	⋯	x_{2p3}	
	x_{313}	x_{323}	x_{333}	⋯	x_{3p3}	
	⋮	⋮	⋮	⋯	⋮	
	$x_{n_3 13}$	$x_{n_3 23}$	$x_{n_3 33}$	⋮	$x_{n_3 p3}$	
Rataan	$\bar{x}_{.13}$	$\bar{x}_{.23}$	$\bar{x}_{.33}$	⋯	$\bar{x}_{.p3}$	$\bar{x}_{..3}$
	⋮	⋮	⋮	⋮	⋮	
g	x_{11g}	x_{12g}	x_{13g}	⋯	x_{1pg}	
	x_{21g}	x_{22g}	x_{23g}	⋯	x_{2pg}	
	x_{31g}	x_{32g}	x_{33g}	⋯	x_{3pg}	
	⋮	⋮	⋮	⋮	⋮	
	$x_{n_i 1g}$	$x_{n_i 2g}$	$x_{n_i 3g}$	⋯	$x_{n_i pg}$	
$\bar{x}_{..g}$	$\bar{x}_{.1g}$	$\bar{x}_{.2g}$	$\bar{x}_{.3g}$	⋯	$\bar{x}_{.pg}$	$\bar{x}_{..g}$
	$\bar{x}_{.1.}$	$\bar{x}_{.2.}$	$\bar{x}_{.3.}$		$\bar{x}_{.p.}$	
Rataan Umum	$\bar{x}_{...}$					

Salah satu bentuk analisis diskriminan yaitu analisis diskriminan linier. Analisis diskriminan linier adalah salah satu metode dalam statistika yang dapat digunakan untuk mengevaluasi klasifikasi objek. Tujuan dari analisis diskriminan linier adalah untuk mendapatkan suatu fungsi yang disebut dengan fungsi diskriminan, dimana fungsi ini dapat memisahkan objek sesuai dengan kelompoknya. Fungsi diskriminan ini juga dapat digunakan untuk memprediksi kelompok dari objek baru yang diamati (Sharma, 1996).

Analisis diskriminan linier menggunakan asumsi data berdistribusi normal multivariat, vektor rata-rata dan kesamaan matriks variansi-kovariansi. Analisis

diskriminan linier memberikan tingkat kesalahan klasifikasi terkecil untuk semua ukuran sampelnya dan analisis diskriminan linier baik digunakan jika matriks variansi-kovariansi sama (Pepler et al, 2016). Asumsi yang harus dipenuhi dalam analisis diskriminan linier adalah data harus berdistribusi normal multivariate, vektor rata-rata yang berbeda, dan matriks variansi-kovariansi harus sama. Bentuk umum fungsi diskriminan linier untuk kasus dua kelompok adalah sebagai berikut (Johnson dan Dean, 2002) :

$$Y = (\bar{X}_1 - \bar{X}_2)' S_G^{-1} X$$

dengan:

\bar{X}_1 : Rata-rata kelompok pertama

\bar{X}_2 : Rata-rata kelompok ke dua

S_G^{-1} : Invers matriks variansi-kovariansi gabungan

untuk matriks variansi-kovariansi gabungan dapat dituliskan sebagai berikut:

$$S_G = \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2}{n_1 + n_2 - 2}$$

bentuk umum matriks variansi-kovariansi :

$$S_G = \frac{\sum_{k=1}^g (n_k - 1) S_k}{\sum_{k=1}^g (n_k - 1)} \quad (2.10)$$

Adapun bentuk umum fungsi diskriminan linier untuk kasus lebih dari dua kelompok atau lebih sebagai berikut (Johnson dan Winchern, 1998) :

$$Y = a'X = a_1X_1 + a_2X_2 + \dots + a_pX_p \quad (2.11)$$

dengan:

$$a' = (a_1, a_2, \dots, a_p)$$

Analisis diskriminan dapat dibentuk dengan menggunakan pendekatan matriks variansi-kovariansi dalam kelompok (**W**) dan matriks variansi-kovariansi antar kelompok (**B**). fungsi diskriminan dibangun sedemikian sehingga dengan menggunakan skor diskriminan maka keragaman relatif antar kelompok terhadap keragaman relatif dalam kelompok menjadi maksimum. Berikut model vektor p-dimensi dari himpunan **X** adalah:

$$X_{jk} \sim N(\mu_k, \Sigma_k) \quad (k = 1, 2, \dots, g ; j = 1, 2, \dots, p)$$

Misalkan N adalah jumlah kelompok pengamatan, g adalah banyaknya kelompok, dan p adalah banyaknya variabel yang menyebabkan orde matriks variansi-kovariansi $p \times p$. Misalkan pula $\mathbf{X} = [\mathbf{X}_1 \ \mathbf{X}_2 \ \dots \ \mathbf{X}_p]$ adalah vektor variabel acak, $\mathbf{X}_{jk} = (\mathbf{X}_{1jk}, \mathbf{X}_{2jk}, \dots, \mathbf{X}_{n_kjk})'$ vektor nilai pada variabel untuk pengamatan kelompok ke- j dalam kelompok ke- k dan n_k pengamatan dalam kelompok ke- k . Asumsi rata-rata vektor populasi $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \dots, \boldsymbol{\mu}_g$ berbeda secara signifikan dengan matriks variansi-kovariansi sama serta berdistribusi normal multivariat. Jika populasi parameter diketahui, maka rumus klasifikasi menjadi:

$$(\mathbf{X} - \bar{\mathbf{X}}_k)' \mathbf{S}_G^{-1} (\mathbf{X} - \bar{\mathbf{X}}_k), \text{ untuk } k = 1, 2, \dots, g$$

dimana \mathbf{X} adalah vektor nilai untuk sampel klasifikasi (Mardia et al., 1979).

Dalam aplikasi, populasi parametrik vektor rata-rata $\bar{\mathbf{X}}$ dan \mathbf{S} tidak diketahui, sehingga perlu menduga parameter tersebut dengan rumus:

$$\bar{\mathbf{X}}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} \mathbf{X}_{ijk}$$

untuk matriks variansi-kovariansi dalam kelompok (\mathbf{W}) pada Persamaan (2.6). Sedangkan matriks variansi-kovariansi antar kelompok (\mathbf{B}) pada Persamaan (2.7).

Berdasarkan asumsi bahwa matriks variansi-kovariansi setiap populasi sama yaitu $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \dots = \boldsymbol{\Sigma}_g$, maka matriks variansi-kovariansi dapat digabung untuk memperoleh matriks gabungan \mathbf{S}_g sebagai penduga matriks gabungan variansi-kovariansi populasi, jika $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_p)$ vektor variabel acak dan memiliki matriks variansi-kovariansi $\boldsymbol{\Sigma}$ berpangkat penuh (full rank) p , simetris, dan definit positif. Analisis diskriminan untuk kasus g kelompok ($g \geq 2$) yang dilambangkan dengan Y_i merupakan kombinasi linear dari p variabel asal.

Matriks variansi-kovariansi antar kelompok dari Y_i adalah $\mathbf{B}' = \mathbf{a}' \mathbf{B} \mathbf{a}$ dan matriks variansi-kovariansi dalam kelompok dari Y_i adalah $\mathbf{W}' = \mathbf{a}' \mathbf{W} \mathbf{a}$. Selanjutnya λ didefinisikan sebagai keragaman relatif antar kelompok terhadap keragaman dalam kelompok yang biasa juga disebut kriteria diskriminan, yakni:

$$\lambda = \frac{\mathbf{a}' \mathbf{B} \mathbf{a}}{\mathbf{a}' \mathbf{W} \mathbf{a}} \tag{2.12}$$

Akan ditentukan \mathbf{a} sedemikian sehingga λ maksimum dan ragam Y_i maksimum ke- i dan Y_i tidak berkorelasi dengan $(i - 1)$ fungsi diskriminan

sebelumnya, dengan kendala $\mathbf{a}'\mathbf{W}\mathbf{a} = 1$. Dengan konsep ekstrim pengganda *lagrange* yaitu:

$$\begin{aligned} F(\mathbf{a}, \lambda) &= \frac{\mathbf{a}'\mathbf{B}\mathbf{a}}{\mathbf{a}'\mathbf{W}\mathbf{a}} + \lambda(1 - \mathbf{a}'\mathbf{W}\mathbf{a}) \\ &= \mathbf{a}'\mathbf{B}\mathbf{a} - \lambda(\mathbf{a}'\mathbf{W}\mathbf{a} - 1) \end{aligned}$$

F diturunkan terhadap \mathbf{a} dan λ , lalu masing-masing disamakan dengan nol (syarat eskrim), diperoleh:

$$(\mathbf{B} - \lambda\mathbf{W})\mathbf{a} = 0 \tag{2.13}$$

Akan mempunyai solusi vektor \mathbf{a} tidak trivial, jika $|\mathbf{B} - \lambda\mathbf{W}| = 0$, dimana rank $\mathbf{B} = \min(g - 1, p)$ dan rank $\mathbf{W} = p$, sehingga $|\mathbf{W}| \neq 0$ (\mathbf{W} non-singular), dengan demikian Persamaan (2.13) dapat ditulis:

$$(\mathbf{W}^{-1}\mathbf{B} - \lambda\mathbf{I})\mathbf{a} = 0 \tag{2.14}$$

Persamaan (2.14) akan mempunyai solusi tidak trivial, bila:

$$|\mathbf{W}^{-1}\mathbf{B} - \lambda\mathbf{I}| = 0 \tag{2.15}$$

Persamaan (2.15) disebut persamaan karakteristik, kemudian nilai λ dapat diperoleh, sedangkan vektor eigen \mathbf{a} diperoleh dari Persamaan (2.13) yang bersesuaian dengan λ_i

Hasil Persamaan (2.15) diperoleh λ_i yang merupakan nilai eigen dari matriks $\mathbf{W}^{-1}\mathbf{B}$, sedangkan \mathbf{a} adalah vektor eigen yang berpadanan dengan nilai eigen λ_i . Jika solusi Persamaan (2.14) diperoleh dari $\lambda_1 > \lambda_2 > \lambda_3 > \dots > \lambda_s$, dimana $s = \min(g - 1, p)$ yang menunjukkan nilai eigen dari matriks $\mathbf{W}^{-1}\mathbf{B}$ dan $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_s$ vektor eigen yang berpadan $\lambda_1, \lambda_2, \dots, \lambda_s$ yang memenuhi kendala $\mathbf{a}^{-1}\mathbf{W}\mathbf{a} = 1$, maka vektor \mathbf{a}_1 merupakan koefisien pembobot fungsi diskriminan ke-1 yang menerangkan keragaman relatif terbesar, vektor eigen \mathbf{a}_2 merupakan koefisien pembobot fungsi diskriminan ke-2 yang menerangkan keragaman relatif sisa terbesar setelah diterangkan oleh fungsi diskriminan ke-1 dan seterusnya (Johnson dan Winchern, 1998).

2.4 Uji Asumsi

Menurut Johnson & Wichern (2007) asumsi data berdistribusi normal, asumsi uji vector rata-rata, dan asumsi kesamaan matriks variansi-kovariansi terpenuhi mengarah ke analisis diskriminan linier. Asumsi data berdistribusi normal

menggunakan jarak Mahalanobis, untuk uji vektor rata-rata menggunakan uji *V-Bartlett*, sedangkan untuk uji kesamaan matriks variansi-kovariansi menggunakan uji *Box's M*.

2.4.1 Distribusi Normal Multivariat

Distribusi normal multivariat merupakan suatu bentuk distribusi data statistika yang saling berhubungan dimana variabel-variabel dari data yang bersangkutan lebih dari dua data, dengan masing-masing variabel memenuhi sifat normalitas. Asumsi distribusi normal multivariat diperlukan untuk menguji signifikan variabel penciri (diskriminator) dan fungsi diskriminan. Pengujian ini dilakukan dengan mengukur jarak Mahalanobis (d_i^2) menggunakan Q-Q plot dari nilai d_i^2 . Hipotesis yang digunakan dalam pengujian distribusi normal multivariat yaitu:

H_0 : data berdistribusi normal multivariat

H_1 : data tidak berdistribusi normal multivariat

Untuk melakukan perhitungan jarak Mahalanobis (d_i^2) diperlukan matriks variansi-kovariansi sebagai berikut:

$$S = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})'$$

dan jarak Mahalanobis didefinisikan sebagai berikut:

$$d_i^2 = (X_i - \bar{X})S^{-1}(X_i - \bar{X})' \quad (2.3)$$

dengan :

d_i^2 : Jarak Mahalanobis pengamatan ke- i

X_i : Pengamatan ke- i

\bar{X} : Rata-rata variabel

S : Matriks variansi-kovariansi

n : Jumlah pengamatan

p : Jumlah variabel prediktor

Apabila titik kordinat di plot jarak Mahalanobis dan chi-kuadrat membentuk garis lurus dan terdapat minimal 50% nilai $d_i^2 \leq \chi_p^2 \left(\frac{i-0,5}{n} \right)$ maka data tersebut dapat dikatakan berdistribusi normal multivariat atau dengan kata lain H_0 diterima (Johnson dan Wichern, 2007).

2.4.2 Uji Vektor Rata-rata

Pada dasarnya fungsi analisis diskriminan digunakan untuk menerangkan perbedaan antar populasi, sehingga hal tersebut dapat dilihat berdasarkan nilai rata-rata populasi yang berbeda. Jika vektor rata-rata antar variabel yang diperoleh sama, maka akibatnya fungsi diskriminan yang diperoleh tidak signifikan. Hipotesis untuk pengujian vektor rata-rata antar kelompok dari sejumlah g kelompok adalah:

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_g = \mu$$

$$H_1 : \mu_i \neq \mu_j \text{ untuk suatu } i \neq j ; i = 1, 2, \dots, g ; j = 1, 2, \dots, g$$

Statistik uji yang digunakan adalah statistik *V-Bartlett*, yang mengikuti sebaran Chi-kuadrat dengan derajat bebas $p(g - 1)$. Statistik *V-Bartlett* dirumuskan sebagai berikut:

$$V = - \left[N - 1 - \frac{(p + g)}{2} \right] \ln \Lambda^* \tag{2.4}$$

dengan :

N : jumlah pengamatan

p : jumlah variabel prediktor

g : jumlah kelompok

Λ^* : Wilk's lambda

$$\Lambda^* = \frac{|W|}{|B + W|} \tag{2.5}$$

dimana:

$$W = \sum_{i=1}^{n_k} \sum_{k=1}^g (X_{ik} - \bar{X}_k)(X_{ik} - \bar{X}_k)' \tag{2.6}$$

$$B = \sum_{k=1}^g (\bar{X}_k - \bar{X})(\bar{X}_k - \bar{X})' \tag{2.7}$$

dengan :

W : Matriks jumlah kuadrat dan hasil kali data dalam kelompok

B : Matriks jumlah kuadrat dan hasil kali data antar kelompok

n_k : Jumlah pengamatan pada kelompok ke-k

X_{ik} : Pengamatan ke-I pada kelompok ke-k

\bar{X}_k : Vektor rata-rata kelompok ke-k

\bar{X} : Vektor rata-rata total

H_0 ditolak jika uji *V-Bartlett* $> \chi^2_{\alpha,p(g-1)}$, atau jika nilai signifikansi lebih dari 0.05, dengan kata lain terdapat perbedaan vektor rata-rata antar kelompok. Sehingga fungsi diskriminan dapat digunakan untuk mengklasifikasikan objek baru ke dalam salah satu kelompok yang telah diketahui (Johnson dan Wichern, 2007).

2.4.3 Uji Kesamaan Matriks Variansi-kovariansi

Kesamaan matriks variansi-kovariansi merupakan salah satu asumsi yang harus dipenuhi pada analisis diskriminan linier. Uji kesamaan matriks variansi-kovariansi menggunakan uji *Box's M*, dengan hipotesis sebagai berikut (Lemeshko dan Mirkin, 2004).

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

$$H_1 : \Sigma_i \neq \Sigma_j \text{ untuk suatu } i \neq j$$

Misalkan S_k adalah penduga tak bias dari Σ_k yang merupakan matriks variansi-kovariansi untuk kelompok ke- k dengan derajat bebas $(n_k - 1)$. S_k diperoleh dari persamaan sebagai berikut:

$$S_k = \frac{1}{n_k - 1} \sum_{i=1}^{n_k} (X_{ki} - \bar{X}_k)(X_{ki} - \bar{X}_k)' \tag{2.8}$$

maka didapatkan persamaan matriks variansi-kovariansi sebagai berikut:

$$S_g = \frac{1}{\sum_{k=1}^g n_k - 1} \sum_{k=1}^g (n_k - 1) S_k$$

statistik uji *Box's M* sebagai berikut :

$$Box's M = M(1 - C) \tag{2.9}$$

nilai M dan C diperoleh dari persamaan berikut:

$$M = \sum_{k=1}^g (n_k - 1) \ln|S| - \sum_{k=1}^g (n_k - 1) \ln|S_k|$$

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

dengan :

p : Jumlah variabel prediktor

g : Jumlah kelompok

n_k : Jumlah pengamatan kelompok ke- k

Keputusan menolak H_0 jika uji *Box's M* $> \chi_{\alpha, v}^2$, di mana $v = \frac{1}{2}(g-1)p(p-1)$ atau jika nilai signifikansi lebih besar dari 0.05. Keputusan tersebut berarti bahwa matriks variansi-kovariansi antar kelompok tidak berbeda secara signifikan.

2.5 Uji Signifikansi Fungsi Linier

Fungsi diskriminan linier yang telah dibentuk diuji signifikansinya menggunakan uji peranan relatif. Peranan relatif dirumuskan sebagai berikut (Johnson dan Wichern, 2007):

$$\text{Peranan relatif } (Y_m) = \frac{\lambda_m}{\sum_{i=1}^m \lambda_i} \times 100\% \quad ; m \leq p \quad (2.16)$$

dimana, λ merupakan akar ciri yang berhubungan dengan fungsi diskriminan linear. Nilai peranan relatif akan membentuk 100% lebih kecil atau mendekati 0 artinya kelompok sangat berbeda, artinya fungsi diskriminan yang diperoleh bisa menentukan suatu objek atau kasus masuk dalam kelompok yang ada (Supranto, 2004).

2.6 Uji Keakuratan Prediksi Klasifikasi

Untuk menilai klasifikasi secara keseluruhan adalah dengan menentukan tingkat akurasi prediksi dari fungsi diskriminan linier. Untuk menguji keakuratan peridiksi dapat dilakukan dengan menggunakan *Press's Q* Statistik.

H_0 : Pengklasifikasian tidak akurat

H_1 : Pengklasifikasian akurat

Statistkik uji yang digunakan *Press's Q* dirumuskan sebagai berikut:

$$\text{Press's } Q = \frac{[n - (qg)]^2}{n(g-1)} \quad (2.17)$$

dengan:

n : Banyaknya sampel

q : Banyaknyak kelompok yang diklasifikasi secara tepat

k : Banyaknya kelompok

Keputusan menolak H_0 jika *Press's Q* $> \chi_{\alpha, 1}^2$, atau jika nilai signifikansi lebih besar dari 0.05. Keputusan tersebut berarti bahwa akurasi pengklasifikasian atau pengelompokan kelompok tepat berdasarkan ukuran sampel dan jumlah kelompoknya (Rofiq., dkk, 2016).

2.7 Pengklasifikasian Data Baru Berdasarkan Fungsi Diskriminan

Pengklasifikasian data baru berdasarkan fungsi diskriminan linier yang telah diperoleh sebelumnya dengan menggunakan rumus sebagai berikut:

$$\sum_{m=1}^r (Y_m - \bar{Y}_{km})^2 \quad (2.18)$$

$$\sum_{m=1}^r \{a'_m(X - X_k)\}^2 \leq \sum_{m=1}^r \{a'_m(X - X_h)\}^2$$

dengan semua $h \neq k; r \leq s, s = \min(g - 1, p)$

dimana :

Y_m : Skor diskriminan ke- m dari objek (individu pengamatan tertentu)

\bar{Y}_{km} : Rata-rata skor diskriminan ke- m dari kelompok ke- k

a'_m : Vektor koefisien fungsi diskriminan ke- m

X : Vektor data pengamatan dari objek yang akan digolongkan

X_k : Vektor nilai rata-rata variabel pembeda dari kelompok ke- k

s :Jumlah fungsi diskriminan yang mungkin dibentuk dalam analisis diskriminan

r : jumlah fungsi diskriminan yang digunakan dalam penggolongan