

PLAGIAT MERUPAKAN TINDAKAN TIDAK TERPUJI

ANALISIS KELOMPOK

Skripsi

Diajukan Untuk Memenuhi Salah Satu Syarat

Memperoleh Gelar Sarjana Pendidikan

Program Studi Pendidikan Matematika



oleh :

Sugih Winarni

NIM : 981414019

**PROGRAM STUDI PENDIDIKAN MATEMATIKA
JURUSAN PENDIDIKAN MATEMATIKA DAN ILMU PENGETAHUAN ALAM
FAKULTAS KEGURUAN DAN ILMU PENDIDIKAN
UNIVERSITAS SANATA DHARMA
YOGYAKARTA**

2004

ANALISIS KELOMPOK

Oleh:

Sugih Winarni

NIM : 981414019

Telah disetujui oleh:

Pembimbing



Ir. Ig. Aris Dwiatmoko, M.Sc

Tanggal: 19 Maret 2004

PLAGIAT MERUPAKAN TINDAKAN TIDAK TERPUJI

ANALISIS KELOMPOK

Dipersiapkan dan ditulis oleh:

Sugih Winarni

NIM: 981414019

Telah dipertahankan di depan Panitia Penguji

pada tanggal 26 Maret 2004

dan dinyatakan telah memenuhi syarat

Susunan Panitia Penguji

	Nama Lengkap	Tanda Tangan
Ketua	: Drs. A. Atmadi, M.Si	
Sekretaris	: Drs. Th. Sugiarto, MT	
Anggota	: Ir. Ig. Aris Dwiatmoko, M.Sc	
Anggota	: Dr. St. Suwarsono	
Anggota	: M. Andy Rudhito, S.Pd., M.Si	

Yogyakarta, 26 Maret 2004

Fakultas Keguruan dan Ilmu Pendidikan

Universitas Sanata Dharma




M. Samet Soewandi, M.Pd.

“... dalam kecemasanku, Engkaulah Tuhan
sumber Pengharapanku ... ”

Buah karya ini ku persembahkan untuk,

♥ *Tuhan Yesus dan Bunda Maria*

♥ *Kedua Orang Tua ku Bp/Ibu Djuri Wijatno*

♥ *Sugih Nugroho dan Sugih Mamungku*

♥ *Prasetyo Wibowo*

PLAGIAT MERUPAKAN TINDAKAN TIDAK TERPUJI

PERNYATAAN KEASLIAN KARYA

Saya menyatakan dengan sesungguhnya bahwa skripsi yang saya tulis ini tidak memuat karya atau bagian karya orang lain, kecuali yang telah disebutkan dalam kutipan dan daftar pustaka, sebagaimana layaknya karya ilmiah.

Yogyakarta, 26 Maret 2004

Penulis



Sugih Winarni



PLAGIAT MERUPAKAN TINDAKAN TIDAK TERPUJI

ABSTRAK

Analisis kelompok adalah salah satu teknik analisis multivariat yang tujuan utamanya mengelompokkan obyek - obyek berdasarkan karakteristik yang dimiliki. Analisis kelompok dilakukan dengan dengan langkah – langkah tertentu, pertama adalah mengukur similaritas antar obyek, proses standarisasi, memilih algoritma pengelompokan, interpretasi kelompok dan diakhiri dengan membuat profil kelompok.

Ada dua metode yang digunakan dalam analisis kelompok, yaitu: metode hierarki dan metode non hierarki. Dalam metode hierarki, pengelompokan dimulai dengan mengukur dua atau lebih obyek yang memiliki derajat similaritas paling tinggi kemudian dilanjutkan dengan obyek lain yang mempunyai derajat similaritas tertinggi kedua. Berbeda dengan metode nonhierarki yang tidak meliputi proses pengelompokan bertingkat seperti metode hierarki. Metode nonhierarki menentukan obyek – obyek yang masuk dalam kelompok dan menentukan banyaknya kelompok yang akan dibentuk.

Setelah dilakukan analisis data dengan menggunakan analisis kelompok, diharapkan terbentuk kelompok dengan homogenitas (similaritas) yang tinggi antar anggota dalam satu kelompok dan heterogenitas (perbedaan) yang tinggi antar kelompok yang satu dengan kelompok yang lain.

ABSTRACT

Cluster analysis is a multivariate analysis technique whose primary purpose is to group objects based on the characteristic they possess. Cluster analysis is expected to form a group with high similarity between the members of the group and high difference between one group and another.

Cluster analysis is done with specific steps there are similarity measures between objects to be clustered, standardizing the data, clustering algorithm, interpretation of the cluster, and profiling the clusters.

Most commonly used clustering algorithms can be classified into two general categories: hierarchical and non - hierarchical. In hierarchical method, each object starts out as its own cluster. In subsequent steps, the two closest clusters are combined into a new aggregate cluster, thus reducing the number of clusters by one in each step. In some cases, a third object join the first two in a cluster. In other, two groups of objects formed at an earlier stage may join together in a new cluster. Eventually, all objects are grouped into one large cluster. In contrast to hierarchical methods, nonhierarchical procedures do not involve the treelike construction process. Instead, they assign objects into clusters once the number of clusters to be formed is specified.

PLAGIAT MERUPAKAN TINDAKAN TIDAK TERPUJI

KATA PENGANTAR

Puji syukur pada Tuhan atas kasih dan penyertaan-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “ Analisis Kelompok “ dengan baik.

Banyak kesulitan dan hambatan yang penulis alami selama proses penyusunan skripsi ini. Namun atas dukungan berbagai pihak penulis dapat menyelesaikan skripsi ini dengan baik. Oleh karena itu, pada kesempatan yang baik ini penulis dengan tulus hati mengucapkan banyak terima kasih kepada:

1. Ir. Ig. Aris Dwiatmoko M. Sc., selaku dosen pembimbing yang dengan penuh kesabaran dan perhatian memberi dorongan dan bimbingan selama proses penyusunan skripsi ini.
2. Drs. Th. Sugiarto MT., selaku Kaprodi atas dukungannya selama ini.
3. Segenap dosen JP. MIPA dan F.MIPA atas ilmu yang diberikan.
4. Bapak Sunarjo dan Bapak Sugeng, atas keramahannya dalam melayani mahasiswa.
5. Kedua orangtuaku Bapak/Ibu Djuri Wijatno atas cinta, doa, dorongan dan kepercayaan yang diberikan.
6. Adikku Sugih Nugroho dan Sugih Manungku, atas perhatian dan cintanya.
7. Prasetyo Wibowo atas cinta, dukungan semangat dan perhatiannya.
8. Teman – teman mahasiswa P. Mat angkatan ‘98’ atas kebersamaannya selama kuliah.

PLAGIAT MERUPAKAN TINDAKAN TIDAK TERPUJI

9. Mince, Zuzan, Rini, Adeck, Diana, Ratna, atas dukungan dan semangatnya.
10. Teman – teman mudika ‘St. Georgeus’ makasih atas dukungan dan semangatnya.
11. Dan semua pihak yang belum disebutkan.

Penulis menyadari bahwapenulisan skripsi ini masih jauh dari sempurna, untuk itu segala saran dan kritik dari berbagai pihak sangat penulis harapkan guna penyempurnaan skripsi ini. Akhirnya penulis berharap semoga skripsi ini bermanfaat bagi para pembaca.

Yogyakarta, Maret 2004

Penulis

PLAGIAT MERUPAKAN TINDAKAN TIDAK TERPUJI

DAFTAR ISI

	Halaman
HALAMAN JUDUL.....	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERSEMBAHAN.....	iv
HALAMAN KEASLIAN KARYA.....	v
ABSTRAK.....	vi
ABSTRACT.....	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR.....	xv
DAFTAR DENDOGRAM.....	xvi
DAFTAR LAMPIRAN.....	xvii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Perumusan Masalah.....	3
1.3 Tujuan Penulisan.....	3
1.4 Manfaat Penulisan.....	3
1.5 Metode Penulisan.....	4
1.6 Pembatasan Masalah.....	4
1.7 Sistematika Penulisan.....	5

PLAGIAT MERUPAKAN TINDAKAN TIDAK TERPUJI

BAB II	MATERI PRASYARAT.....	6
2.1	Variabel.....	6
2.2	Pengukuran dan Hasil pengukuran.....	6
2.2.1	Skala Nominal.....	12
2.2.2	Skala Ordinal.....	13
2.2.3	Skala Interval.....	14
2.2.4	Skala Rasio.....	14
2.3	Ukuran Derajat Similaritas.....	15
2.3.1.	Jarak Antara Dua Obyek.....	16
2.3.2.	Koefisien Assosiasi Antar Obyek.....	18
2.3.3.	Koefisien Korelasi.....	23
2.4	Analisis Variansi (ANOVA).....	32
BAB III	ANALISIS KELOMPOK.....	35
3.1.	Pendahuluan.....	35
3.2.	Proses Dasar Analisis Kelompok.....	37
3.2.1	Mengukur Similaritas Antar Obyek.....	38
3.2.2	Proses Standarisasi.....	38
3.2.3	Memilih Algoritma Pengelompokan.....	44
3.2.4	Interpretasi Kelompok.....	85
3.2.5	Validasi dan Membuat profil.....	85
BAB IV	APLIKASI ANALISIS KELOMPOK.....	92
4.1.	Dalam Bidang Pendidikan.....	92
4.2	Dalam Bidang Kesehatan.....	100

PLAGIAT MERUPAKAN TINDAKAN TIDAK TERPUJI

BAB V PENUTUP.....	109
5.1 Kesimpulan	109
DAFTAR PUSTAKA.....	xviii



PLAGIAT MERUPAKAN TINDAKAN TIDAK TERPUJI

DAFTAR TABEL

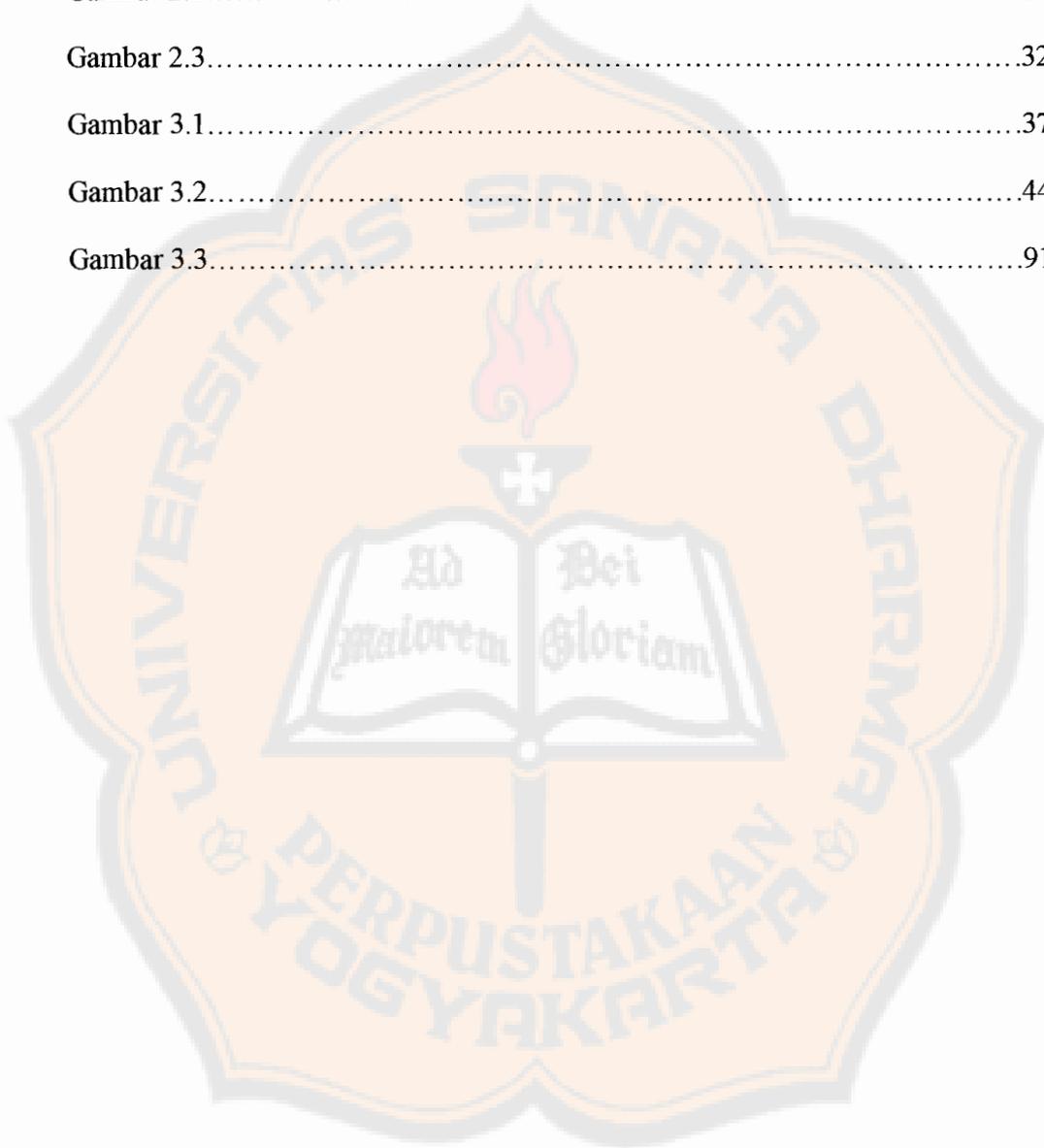
	Halaman
Tabel 2.1.....	12
Tabel 2.2.....	18
Tabel 2.3.....	19
Tabel 2.4.....	20
Tabel 2.5.....	21
Tabel 2.6.....	23
Tabel 2.7.....	34
Tabel 2.8.....	34
Tabel 3.1.....	41
Tabel 3.2.....	43
Tabel 3.3.....	64
Tabel 3.4.....	65
Tabel 3.5.....	66
Tabel 3.6.....	72
Tabel 3.7.....	72
Tabel 3.8.....	73
Tabel 3.9.....	74
Tabel 3.10.....	75
Tabel 3.11.....	76
Tabel 3.12.....	79

PLAGIAT MERUPAKAN TINDAKAN TIDAK TERPUJI

Tabel 3.13.....	79
Tabel 3.14.....	80
Tabel 3.15.....	81
Tabel 3.16.....	81
Tabel 3.17.....	82
Tabel 3.18.....	83
Tabel 3.19.....	83
Tabel 3.20.....	86
Tabel 4.1.....	95
Tabel 4.2.....	95
Tabel 4.3.....	96
Tabel 4.4.....	98
Tabel 4.5.....	98
Tabel 4.6.....	98
Tabel 4.7.....	99
Tabel 4.8.....	103
Tabel 4.9.....	103
Tabel 4.10.....	104
Tabel 4.11.....	106
Tabel 4.12.....	106
Tabel 4.13.....	107
Tabel 4.14.....	107

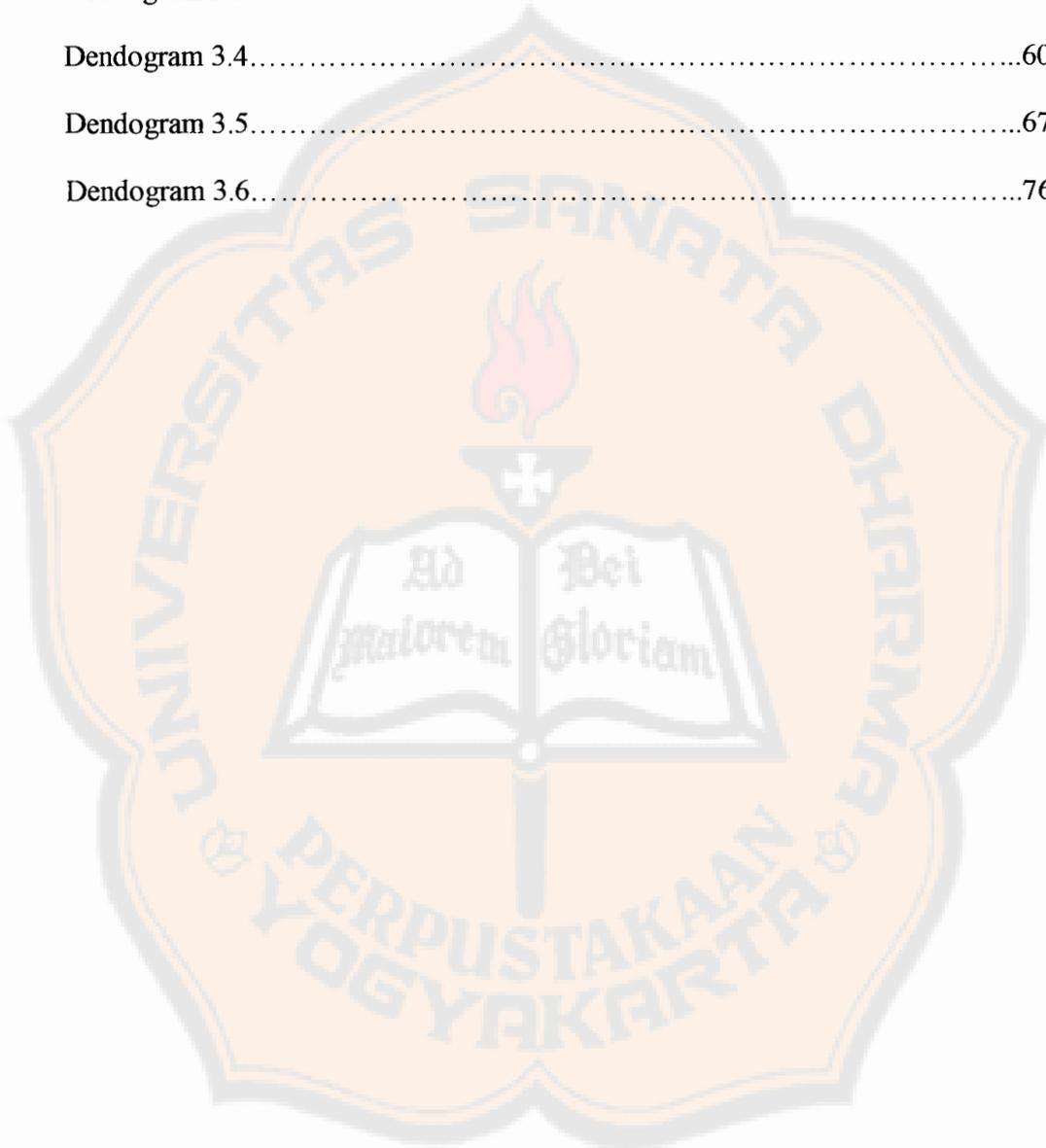
DAFTAR GAMBAR

	Halaman
Gambar 2.1.....	7
Gambar 2.2.....	14
Gambar 2.3.....	32
Gambar 3.1.....	37
Gambar 3.2.....	44
Gambar 3.3.....	91



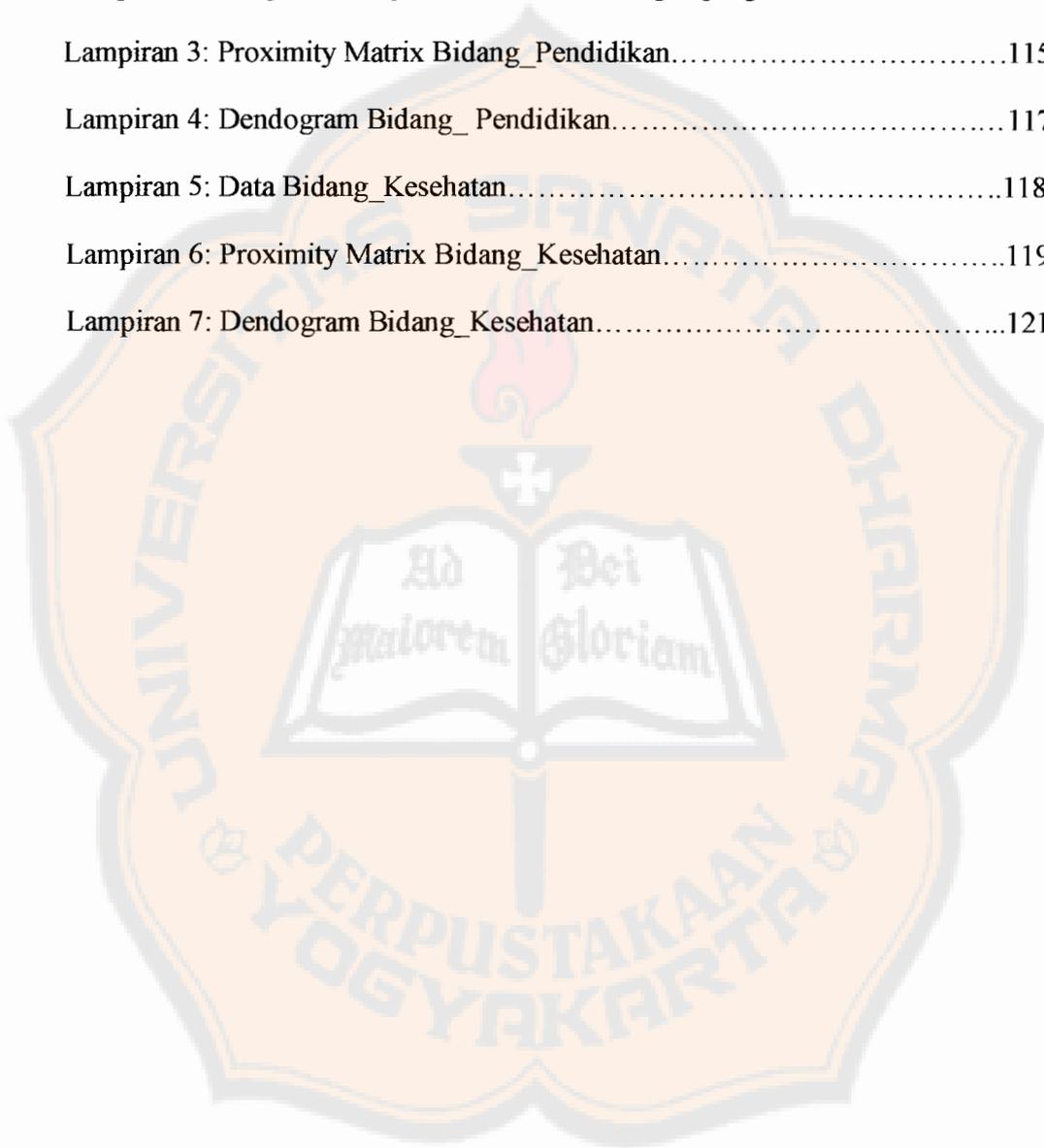
DAFTAR DENDOGRAM

	Halaman
Dendogram 3.1.....	47
Dendogram 3.2.....	50
Dendogram 3.3.....	52
Dendogram 3.4.....	60
Dendogram 3.5.....	67
Dendogram 3.6.....	76



DAFTAR LAMPIRAN

	Halaman
Lampiran 1: Data Bidang_Pendidikan.....	111
Lampiran 2: Langkah – langkah analisis data dengan program SPSS.....	112
Lampiran 3: Proximity Matrix Bidang_Pendidikan.....	115
Lampiran 4: Dendogram Bidang_Pendidikan.....	117
Lampiran 5: Data Bidang_Kesehatan.....	118
Lampiran 6: Proximity Matrix Bidang_Kesehatan.....	119
Lampiran 7: Dendogram Bidang_Kesehatan.....	121



PLAGIAT MERUPAKAN TINDAKAN TIDAK TERPUJI

BAB I

PENDAHULUAN

1.1 LATAR BELAKANG

Analisis kelompok merupakan salah satu metode analisis data yang penting untuk mengelompokkan data berdasar karakteristik yang dimiliki. Dalam kehidupan sehari – hari banyak dijumpai masalah – masalah yang memerlukan penyelesaian. Misalnya ingin diketahui perilaku konsumen yang akan membeli rumah di lingkungan real estate mempunyai beberapa hal yang harus dipertimbangkan seperti lingkungan yang aman, lingkungan yang nyaman dan bebas banjir, fasilitas umum dan sosial yang memadai, letak perumahan yang strategis, dekat dengan pasar dan sekolah, harga terjangkau, sistem pembayaran yang fleksibel, dan lain sebagainya. Masalah ini memang rumit karena mencakup banyak kriteria yang menentukan hasil dari pengujian.

Dalam statistika, masalah diatas dapat di analisis dengan analisis multivariat. Data diatas disebut data multivariat yaitu data yang diperoleh dengan mengukur lebih dari satu variabel kriteria pada setiap individu. Analisis multivariat adalah teknik – teknik analisis statistik yang memperlakukan sekelompok variabel kriteria yang saling berkorelasi sebagai satu sistem, dengan memperhitungkan korelasi antar variabel – variabel itu.

Salah satu tujuan dari analisis multivariat adalah menemukan dan menafsirkan struktur atau ciri yang mendasari data. Agar struktur atau ciri yang

ditemukan itu bermakna maka peneliti perlu bijaksana dalam memilih variabel kriteria mana yang diukur.

Ada beberapa teknik analisis multivariat yang dapat digunakan untuk menyelesaikan masalah diatas. Salah satunya adalah dengan analisis kelompok. Analisis kelompok adalah teknik pengelompokan yang mengelompokan obyek berdasarkan karakteristik yang mereka miliki. Jadi setiap obyek yang memiliki kesamaan dengan obyek yang lain akan menjadi satu kelompok. Hasil pengelompokan obyek ini seharusnya dapat menunjukkan homogenitas kedalam (dalam satu kelompok) dan heterogenitas keluar (antar kelompok).

Fokus utama analisis kelompok adalah kesamaan obyek berdasar pada variat dan bukan pada taksiran variat itu sendiri. Hal ini menjadikan pendefinisian variat oleh peneliti sebagai suatu langkah penting dalam analisis kelompok. Kelompok variat (*Cluster Variate*) adalah himpunan variabel yang mewakili karakteristik yang digunakan untuk membandingkan obyek dalam analisis kelompok. Oleh karena itu, analisis variat menentukan 'karakter' obyek tersebut.

Data perlu dikelompokan dengan analisis kelompok karena dengan analisis kelompok dapat memberikan arti informal dalam menaksir data, mengidentifikasi data yang keluar, dan memberikan hipotesis menarik yang menyatakan hubungan.

Analisis kelompok biasanya melibatkan paling sedikit tiga langkah yaitu

1. Mengukur beberapa similaritas yang ada dalam obyek sehingga dapat digunakan untuk menentukan berapa banyak kelompok yang ada dalam sampel.
2. Proses pengelompokan yang nyata dimana setiap obyek dibagi kedalam beberapa kelompok
3. Menunjukkan setiap obyek masuk dalam kelompok yang mana.

Analisis kelompok digunakan dalam beberapa bidang. Misalnya: dalam bidang Psikologi, Ekonomi, Pendidikan, Sosiologi, Geologi, Biologi, Meteorologi dan masih banyak lagi. Dalam penulisan skripsi ini, akan dibahas penerapan analisis kelompok dalam bidang pendidikan dan bidang kesehatan.

1.2 PERUMUSAN MASALAH

Adapun masalah yang dibahas dalam skripsi ini adalah:

1. Apa yang dimaksud dengan Analisis Kelompok ?
2. Bagaimana prosedur Analisis Kelompok ?
3. Bagaimanⁿ penerapan Analisis Kelompok untuk analisis data?

1.3 TUJUAN PENULISAN

Tujuan penulisan skripsi ini adalah membahas dan menerapkan Analisis Kelompok sebagai salah satu cara dalam menganalisis data multivariat.

1.4 MANFAAT PENULISAN

Dalam mempelajari Analisis Kelompok, penulis mempunyai harapan semoga penulisan skripsi ini memberikan manfaat bagi berbagai pihak terutama:

- Bagi Universitas Sanata Dharma

Diharapkan dapat menambah referensi kepustakaan dan dapat digunakan sebagai bahan bagi pihak – pihak yang memerlukannya.

- Bagi peneliti

Dapat dijadikan sebagai bahan pertimbangan dalam melakukan analisis data terutama untuk data multivariat.

- Bagi Penulis

Penulis mendapat pengetahuan, pengalaman dan wawasan yang luas mengenai analisis kelompok sebagai salah satu teknik dalam menganalisis data multivariat.

1.5 METODE PENULISAN

Penulisan skripsi ini menggunakan metode studi pustaka yaitu dengan mempelajari buku – buku pendukung yang berkaitan dengan analisis kelompok, sehingga dalam penulisan ini tidak ditemukan hal – hal yang baru.

1.6 PEMBATASAN MASALAH

Analisis Kelompok dibahas secara deskriptif. Beberapa topik analisis kelompok yang tidak dibahas dalam penulisan skripsi ini adalah sebagai berikut:

1. Beberapa metode dalam pengelompokan Nonhierarki

2. Matriks sebagai materi prasyarat tidak dibahas secara rinci.
3. Penentuan jumlah optimal kelompok.

1.7 SISTEMATIKA PENULISAN

Bab I Pendahuluan

Berisi Latar Belakang, Perumusan Masalah, Tujuan Penulisan, Manfaat Penulisan, Metode Penulisan, Pembatasan Masalah, dan Sistematika Penulisan.

Bab II Materi Prasyarat

Dibahas mengenai materi – materi yang menjadi dasar untuk mempelajari analisis kelompok yaitu: Tipe Data Statistik dan Ukuran Derajat Kesamaan.

Bab III Analisis Kelompok

Membahas analisis kelompok dan proses dasar dalam analisis kelompok.

Bab IV Aplikasi

Membahas aplikasi analisis kelompok dalam berbagai bidang aplikasi.

Bab V Penutup

Berisi kesimpulan.

PLAGIAT MERUPAKAN TINDAKAN TIDAK TERPUJI

BAB II

MATERI PRASYARAT

2.1. Variabel

Dalam subbab ini akan dibahas mengenai variabel yang sangat penting dalam suatu penelitian. Variabel harus ditetapkan terlebih dahulu sebelum melakukan proses pengukuran.

Definisi 2.1.1

Variabel adalah suatu karakteristik yang dimiliki oleh suatu obyek yang membedakannya dengan yang lain.

Sebagai contoh, obyek – obyek seperti orang, binatang atau benda mungkin memiliki beberapa karakteristik seperti berat, ukuran, umur, warna mata dan sebagainya. Contoh variabel yang penting dalam bidang sosiologi, psikologi, dan pendidikan adalah: jenis kelamin, penghasilan, kelas sosial, produktifitas organisasi, mobilitas pekerjaan, tingkat aspirasi, kecakapan verbal, kecemasan, daya ingat, dan prestasi.

2.2. Pengukuran dan Hasil Pengukuran

Tujuan suatu proses pengukuran, seperti pengumpulan data adalah menggolongkan obyek ke dalam kategori variabel. Masalah pengukuran memegang peranan yang sangat penting dalam penelitian, terutama dalam penelitian ilmu – ilmu sosial karena variabel dalam ilmu sosial berasal dari suatu konsep yang perlu diperjelas dan diubah bentuknya sehingga dapat diukur.

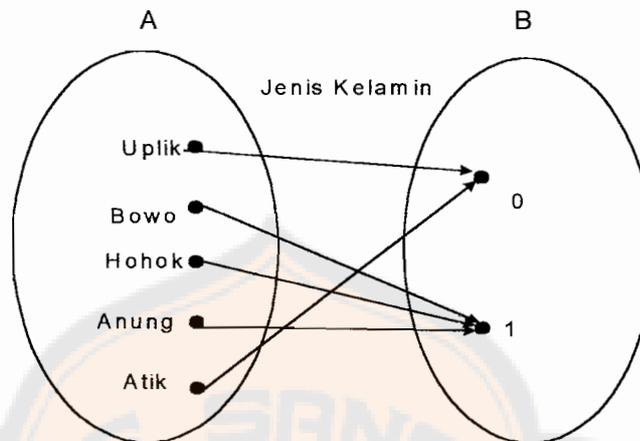
Pengukuran variabel tertentu kadang sulit dilakukan. Pengukuran variabel jenis kelamin atau kelas sosial lebih mudah dilakukan daripada pengukuran variabel kreativitas, konservatisme, atau efektivitas organisasi. Sebagai contoh: variabel kecemasan didefinisikan sebagai rasa takut yang subyektif. Untuk mempermudah pengukuran dan menganalisis data maka variabel kecemasan yang merupakan variabel kontinu harus diubah menjadi variabel dikotomis misalnya 'cemas dan tidak cemas' atau variabel politomis misalnya 'tingkat kecemasan rendah, sedang, dan tinggi'.

Definisi 2.2.1

Pengukuran adalah pemberian nilai / bilangan pada obyek – obyek berdasarkan pada variabel yang telah ditetapkan. (Stevens, 1951 dalam Kerlingers, 1990)

'Pemberian' dalam definisi diatas merupakan pemetaan obyek dari satu himpunan pada obyek dari himpunan lain. Fungsi adalah aturan yang memetakan setiap anggota suatu himpunan pada setiap anggota himpunan lain. Jadi semua prosedur pengukuran merupakan suatu fungsi, dengan obyek – obyek yang diukur dipandang sebagai domain dan besaran yang dipetakan pada obyek tersebut diipandang sebagai kodomain.

Sebagai contoh, untuk variabel jenis kelamin jika jenis kelamin laki – laki diberi nilai 1 dan jika jenis kelamin perempuan diberi nilai 0. Domain variabel adalah himpunan anak, misalnya himpunan $A = \{\text{Bowo, Anung, Atik, Uplik, Hohok}\}$. Kodomain variabel adalah himpunan nilai, misalnya himpunan $B = \{0, 1\}$. Variabel jenis kelamin ini dapat digambarkan dalam pemetaan berikut:



Gambar 2.1

Ada dua macam data hasil pengukuran yang akan dibicarakan dalam skripsi ini, yaitu data kuantitatif (metrik) dan data kualitatif (non metrik).

Data Kualitatif (non metrik) terdiri atas bilangan – bilangan yang menyatakan sifat – sifat, karakteristik, atau kategori yang mengidentifikasi / menggambarkan suatu obyek. Data non metrik menggambarkan perbedaan (ada atau tidaknya) tipe atau macam karakteristik yang dimiliki suatu obyek. Sebagai contoh, jika suatu obyek dinyatakan laki – laki maka obyek tersebut tidak dapat dinyatakan sebagai perempuan.

Data Kuantitatif (metrik) adalah data yang membedakan obyek berdasarkan perbedaan besar / kuantitas atau derajat karakteristik yang diukur. Sebagai contoh, umur dan berat badan seseorang merupakan data metrik.

Dalam statistik, istilah skala pengukuran digunakan untuk menggambarkan proses sistematis dengan memberikan bilangan atau nama pada obyek dan atribut yang dimiliki. Sebagai contoh, orang dapat dilihat sebagai

sebuah obyek. Nama yang diberikan pada orang adalah simbol yang digunakan untuk menunjukkan obyek tersebut.

Ketika dilakukan perhitungan pada bilangan yang menunjukkan obyek atau sifat yang dimiliki, baik dengan penjumlahan, pengurangan, pembagian, perkalian atau mencari rata – rata, harus diperhatikan karakteristik obyek yang ditunjukkan dengan bilangan. Sebagai contoh, rata – rata dua nomor rumah atau dua nomor telepon tidak relevan untuk dihitung karena tidak memiliki arti.

Dalam statistik, informasi yang dimiliki sebuah obyek dapat diklasifikasikan dalam empat kategori umum, yaitu: skala nominal, skala ordinal, skala interval dan skala rasio. Ada tiga kriteria yang digunakan untuk menyatakan skala pengukuran, yaitu:

1. Besar / Ukuran.

Kriteria besar / ukuran memungkinkan dilakukan proses perhitungan atau pengukuran. Sebagai hasilnya, dapat dibandingkan karakteristik antara satu obyek dengan obyek yang lain menggunakan kata “lebih besar dari”, “lebih dari”, atau “sama dengan”.

Tidak semua bilangan yang menyatakan suatu obyek mempunyai syarat ini, karena tidak semua bilangan yang diperoleh merupakan hasil dari proses perhitungan atau pengukuran. Sebagai contoh, tiga nama jalan raya yang diidentifikasi dengan bilangan 78, 80, dan 95. Bilangan – bilangan tersebut tidak diperoleh melalui proses perhitungan atau proses pengukuran. Bilangan – bilangan tersebut tidak memberikan informasi tentang kondisi jalan atau kualitas

jalan raya. Jalan raya yang dinyatakan dengan bilangan 95 tidak lebih baik dari jalan raya yang dinyatakan dengan bilangan 78.

Suatu variabel mempunyai kriteria besar / ukuran jika memberikan informasi tentang posisi karakteristik yang diukur. Sebagai contoh, variabel umur dan variabel bobot mempunyai kriteria besar/ukuran karena nilai variabel tersebut memberikan informasi tentang posisi besar/ukuran karakteristik obyek yang diukur. Sebagai contoh, seseorang dengan umur 45 tahun dikatakan lebih tua dari seseorang dengan umur 25 tahun.

2. Interval Sama

Salah satu cara untuk memahami konsep interval sama adalah dengan menggunakan garis bilangan. Sebagai contoh, diberikan garis bilangan sebagai berikut:



Pada garis bilangan tersebut, nilai dinyatakan dengan setiap tanda interval pada garis. Nilai yang dinyatakan pada setiap tanda interval disebut skala. Setiap tanda mempunyai jarak yang sama dari tanda berikutnya dan bertambah dengan besar yang sama seperti berpindah dari kiri ke kanan pada garis bilangan. Besarnya jarak yang sama dari satu tanda dengan tanda berikutnya disebut unit skala.

Variabel dengan unit skala khusus mempunyai nama khusus seperti inci, detik, mil, gram dan sebagainya. Unit skala harus mempunyai besar yang sama tanpa memperhatikan dimana letak pada skala.

Variabel dengan kriteria interval sama tidak selalu mempunyai arti bila dilakukan operasi penjumlahan dan pengurangan. Sebagai contoh, prestasi dua pekerja, seseorang mendapat peringkat “superior” dan seorang yang lain “dibawah rata – rata”, pada suatu skala kinerja yang menunjukkan peringkat superior, rata – rata dan dibawah rata – rata. Penjumlahan / pengurangan prestasi dua pekerja tersebut tidak ada artinya. Demikian juga, tidak relevan untuk mengetahui seberapa besar jarak prestasi pekerja superior dengan prestasi pekerja dibawah rata – rata. Informasi yang dapat diketahui hanyalah bahwa prestasi superior mempunyai kualitas lebih baik daripada prestasi dibawah rata – rata.

Contoh variabel yang mempunyai kriteria interval sama adalah pendapatan dan tinggi badan seseorang. Bila tinggi badan seseorang 170 cm dan tinggi badan yang lain 159 cm, maka kedua orang tersebut mempunyai selisih tinggi badan 11 cm.

3. Titik Nol Mutlak

Nol mempunyai arti bahwa karakteristik atau kualitas yang diukur tidak ada. Sebagai contoh, tinggi nol artinya tidak ada tinggi, bobot nol artinya tidak berbobot, pendapatan nol artinya tidak mempunyai pendapatan. Dengan kata lain nilai nol bersifat mutlak.

Tidak semua variabel memiliki kriteria titik nol mutlak. Temperatur dalam derajat Celcius tidak mempunyai nol mutlak. Nol derajat tidak mengindikasikan bahwa obyek tidak mempunyai panas. -3 derajat Celcius bukan berarti 3 derajat kurang dari tidak panas.

Secara umum, ketiga kriteria ini digunakan untuk mengklasifikasikan suatu variabel kedalam empat skala pengukuran.

Tabel berikut menunjukkan skala pengukuran dan kriteria yang dimiliki.

Skala	Besar / Ukuran	Interval Sama	Titik Nol Mutlak
Nominal	Tidak	Tidak	Tidak
Ordinal	Ya	Tidak	Tidak
Interval	Ya	Ya	Tidak
Rasio	Ya	Ya	Ya

Tabel 2.1

2.2.1. Skala Nominal

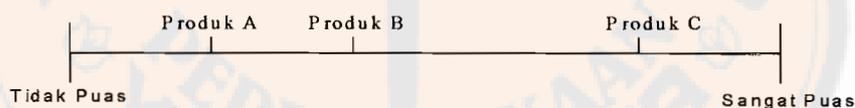
Skala nominal merupakan skala yang paling lemah diantara ke empat skala pengukuran. Skala nominal tidak memiliki ketiga kriteria yang ada. Bilangan yang diberikan pada obyek – obyek merupakan angka yang tidak mengandung arti kuantitatif. Bilangan – bilangan itu tidak dapat diurutkan atau dioperasikan secara matematis.

Skala nominal membedakan obyek yang satu dengan yang lainnya berdasarkan nama atau predikat. Sebagai contoh, bayi yang baru lahir dapat diklasifikasikan menurut variabel jenis kelamin sebagai laki – laki dan perempuan. Pasien di rumah sakit jiwa dapat diklasifikasikan berdasar variabel jenis penyakit sebagai penderita schizophrenic, manic- depressive, psychoneurotic, dan sebagainya. Tidak jarang digunakan nomor – nomor yang di pilih sekehendak hati sebagai pengganti nama – nama untuk membedakan obyek

berdasarkan karakteristik tertentu. Sebagai contoh, digunakan nomor 1 untuk menyebut kelompok barang – barang yang cacat dari suatu proses produksi dan 0 untuk kelompok barang – barang yang tidak cacat. Data semacam ini sering disebut data hitungan (*count data*) atau data frekuensi.

2.2.2. Skala Ordinal

Skala Ordinal mempunyai tingkatan yang lebih tinggi dari pada skala nominal karena memiliki satu kriteria yaitu besar / ukuran. Variabel – variabel dapat diurutkan atau diperingkat dengan skala ordinal berkaitan dengan besarnya atribut yang dimiliki setiap variabel. Setiap sub kelas dapat dibandingkan dengan yang lain dalam hubungan ‘lebih dari’ atau ‘kurang dari’. Sebagai contoh, perbedaan tingkat kepuasan setiap individu terhadap beberapa produk baru dapat digambarkan dalam skala ordinal. Skala berikut menunjukkan pandangan seorang responden terhadap tiga produk. Responden lebih puas dengan produk A dari pada produk B dan lebih puas dengan produk B dari pada produk C.



Gambar 2.2

Bilangan yang digunakan dalam skala ordinal mempunyai arti non kuantitatif. Bilangan hanya mengindikasikan posisi relatif dalam rangkaian terurut. Dalam hal ini tidak ada ukuran kepuasan konsumen secara mutlak dan

peneliti juga tidak mengetahui perbedaan kepuasan responden dengan pasti dalam skala kepuasan. Data semacam ini disebut data peringkat (*rank data*)

2.2.3. Skala Interval

Pada skala interval dimungkinkan untuk membandingkan kuantitas antar obyek dan mengukur besarnya perbedaan antar unit yang berdekatan. Seberapa besar kurang atau lebihnya kualitas obyek yang diukur dibanding obyek yang lain dapat dinyatakan secara numerik.

Data dengan skala interval selalu numerik. Operasi matematis seperti penjumlahan, pengurangan dan mencari rata – rata mungkin dilakukan dan mempunyai arti. Operasi perkalian dan pembagian tidak mungkin dilakukan pada skala interval. Sebagai contoh, temperatur cuaca kemarin adalah $40^{\circ}F$ dan hari ini adalah $65^{\circ}F$. Perbedaan antara temperatur hari ini dan kemarin adalah $25^{\circ}F$. Dengan skala interval, tidak hanya dikatakan temperatur hari ini lebih tinggi dari kemarin, tapi juga dikatakan temperatur hari ini $25^{\circ}F$ lebih tinggi dari kemarin. Tidak ada artinya membagi atau mengalikan temperatur $65^{\circ}F$ dengan $25^{\circ}F$.

2.2.4. Skala Rasio

Sifat – sifat yang dimiliki oleh skala rasio terdapat dalam ketiga skala yang pertama serta sifat tambahan bahwa rasio antara masing – masing pengukuran mempunyai arti. Dalam skala rasio dapat dilakukan semua operasi matematika seperti penjumlahan, pengurangan, perkalian, dan pembagian. Sebagai contoh, pengukuran tinggi dan berat badan seseorang. Seseorang dengan berat 90 kg

memiliki kelebihan berat 30 kg di banding dengan orang yang mempunyai berat 60 kg. Orang dengan berat 90 kg dapat dikatakan dua kali lebih berat dari orang dengan berat 45 kg.

Perbedaan nyata antara skala interval dan skala rasio adalah skala interval tidak mempunyai titik nol mutlak sedangkan skala rasio mempunyai titik nol mutlak. Sebagai contoh, temperatur dengan skala Fahrenheit dan Celcius mempunyai perbedaan titik nol tidak mutlak. Untuk itu, tidak mungkin menyatakan beberapa nilai pada skala interval sebagai kelipatan beberapa nilai yang lain. Tidak benar menyatakan suhu yang terukur dengan temperatur skala $80^{\circ}F$ sebagai dua kali suhu dengan temperatur skala $40^{\circ}F$. Meskipun $80^{\circ}F$ adalah kelipatan dua dari $40^{\circ}F$ tetapi tidak dapat dinyatakan bahwa benda yang suhunya $80^{\circ}F$ mempunyai panas dua kali panas benda yang suhunya $40^{\circ}F$.

Data kualitatif (non metrik) meliputi data dengan skala nominal dan data dengan skala ordinal. Skala ordinal termasuk data kualitatif karena sekalipun ada perbedaan tertentu, skala ordinal tidak mempunyai ukuran yang eksak. Data kuantitatif (metrik) meliputi data dengan skala interval dan data dengan skala rasio.

2.3. Ukuran Derajat Similaritas

Derajat similaritas antar obyek perlu diukur untuk mengetahui seberapa besar similaritas yang dimiliki antar obyek, karena pada dasarnya analisis kelompok mengelompokkan obyek yang mempunyai similaritas. Sebelum membicarakan masalah ukuran derajat similaritas, perlu diketahui jenis data yang

diperoleh berupa data metrik atau data non metrik. Jika jenis datanya metrik, maka ukuran derajat similaritas yang digunakan adalah jarak dan koefisien korelasi. Sedangkan jika data yang diperoleh adalah data non metrik, maka ukuran derajat similaritas yang digunakan adalah koefisien asosiasi.

Konsep similaritas merupakan dasar dari analisis kelompok. Ada tiga ukuran derajat similaritas antar obyek, yaitu:

2.3.1. Jarak Antara Dua Obyek

Ada beberapa macam ukuran jarak. Ukuran jarak digunakan jika jenis datanya metrik. Dalam penulisan skripsi ini, jarak Euclidean yang paling banyak digunakan untuk mengukur similaritas antar obyek, namun tidak ada paksaan untuk menggunakan ukuran tersebut, bahkan dalam situasi tertentu ukuran yang lain lebih memberikan keuntungan. Bila diambil data dengan n obyek, setiap obyek akan diwakili oleh vektor $x' = (x_1, x_2, \dots, x_p)$ pada p variabel. Notasi $x'_i = (x_{i1}, x_{i2}, \dots, x_{ip})$ menyatakan ukuran pada obyek ke - i.

Jarak antara obyek i dan obyek j adalah $d_{x_i x_j}$. Untuk lebih memudahkan, jarak antara obyek i dan obyek j ditulis d_{ij} .

Beberapa ukuran jarak:

- **Jarak Minkowski** didefinisikan oleh: $d_{ij} = \left\{ \sum_{k=1}^p |x_{ik} - x_{jk}|^r \right\}^{\frac{1}{r}}$ dimana

d_{ij} menyatakan jarak antara obyek i dan j dengan r adalah bilangan real ≥ 1

- Jika $r = 2$, diperoleh **Jarak Euclidean** antara obyek i dan j yang

didefinisikan sebagai berikut:
$$d_{ij} = \left\{ \sum_{k=1}^p |x_{ik} - x_{jk}|^2 \right\}^{\frac{1}{2}}$$

- Jika $r = 1$ diperoleh $d_{ik} = \sum_{k=1}^p |x_{ik} - x_{jk}|$ yang disebut sebagai **ukuran mutlak atau ukuran City-Block**.

Jarak Euclidean dan ukuran City – Block memenuhi syarat – syarat matematis dari fungsi jarak sebagai berikut:

1. $d_{ij} \geq 0$ (jarak tak negatif)
2. $d_{ii} = 0$ (jarak obyek dengan dirinya sendiri adalah nol)
3. $d_{ij} = d_{ji}$ (fungsi simetri jarak)
4. $d_{ij} \leq d_{ih} + d_{hj}$ (ketaksamaan segitiga untuk semua i, j , dan h)

$d_{ij} = 0$ tidak berimplikasi $i = j$, karena hal ini dapat terjadi pada 2 obyek yang berbeda dan mempunyai ukuran yang sama untuk variabel yang dipelajari. Misalnya, dua obyek A dan B yang berbeda mempunyai berat yang sama yaitu 130 pon. Tetapi bila $d_{ij} = 0$ maka ketaksamaan segitiga menyatakan bahwa obyek i dan j mempunyai jarak yang sama dengan sembarang obyek yang lain h karena $d_{ih} \leq d_{ij} + d_{jh} = d_{jh}$ sedangkan $d_{jh} \leq d_{ji} + d_{ih} = d_{ih}$, sehingga dapat disimpulkan bahwa $d_{ih} = d_{jh}$

Sebagai contoh, diberikan data tiga obyek dengan berat (pon) dan tinggi (kaki). Hitung jarak Euclidean antar ketiga obyek!

Obyek	Berat (Pon)	Tinggi (Kaki)
A	160	5.5
B	163	6.2
C	165	6

Tabel 2.2

Dapat dihitung jarak Euclidean sebagai berikut:

$$d_{AB} = \sqrt{(160 - 163)^2 + (5.5 - 6.2)^2} = \sqrt{9.49} = 3.08$$

$$d_{AC} = \sqrt{(160 - 165)^2 + (5.5 - 6.0)^2} = \sqrt{25.25} = 5.02$$

$$d_{BC} = \sqrt{(163 - 165)^2 + (6.2 - 6.0)^2} = \sqrt{4.04} = 2.01$$

2.3.2 Koefisien Asosiasi antar obyek

Koefisien asosiasi digunakan untuk mengukur variabel data non metrik (data dengan skala nominal atau data dengan skala ordinal). Sebagai contoh, responden dapat menjawab ‘ya’ atau ‘tidak’ pada beberapa pernyataan yang diberikan.

Koefisien Asosiasi dapat menaksir derajat kesesuaian antar pasangan responden. Bentuk sederhana dari koefisien asosiasi dapat dilihat pada persen kesesuaian (antara responden yang berkata ‘ya’ atau masing – masing berkata ‘tidak’ untuk sebuah pernyataan).

Diberikan tabel kontingensi frekuensi pasangan dan bukan pasangan untuk item i dan item k sebagai berikut :

		item k		total
		1	0	
Item i	1	a	b	a+b
	0	c	d	c+d
total		a+c	b+d	a+b+c+d

Tabel 2.3

a, b, c, dan d adalah frekuensi setiap pasang item yang sesuai pada matriks 2 x 2. Jika kedua item secara serentak bernilai 1 maka frekuensinya dinyatakan dengan a. Jika kedua item secara serentak bernilai 0 maka frekuensinya dinyatakan dengan d. Jika item k bernilai 0 sedang item i bernilai 1 maka frekuensinya dinyatakan dengan b. Jika item k bernilai 1 sedang item i bernilai 0 maka frekuensinya dinyatakan dengan c.

Dari tabel, didapatkan koefisien asosiasi untuk mengelompokkan item – item:

No	Koefisien	Alasan
1	$\frac{a+d}{a+b+c+d}$	Kesamaan Bobot untuk 1 - 1 dan 0 - 0 (Pasangan yang cocok)
2	$\frac{2(a+d)}{2(a+d)+b+c}$	Bobot 2 untuk 1 -1 dan 0 - 0 (pasangan yang cocok)
3	$\frac{a+d}{a+d+2(b+c)}$	Bobot 2 untuk pasangan yang tidak cocok (1 - 0 atau 0 - 1)
4	$\frac{a}{a+b+c+d}$	Tidak ada pasangan 0 - 0 dalam pembilang
5	$\frac{a}{a+b+c}$	Tidak ada pasangan 0 - 0 dalam pembilang atau penyebut

6	$\frac{2a}{2a+b+c}$	Tidak ada pasangan 0 - 0 sebagai pembilang atau penyebut. Bobot 2 untuk pasangan 1
7	$\frac{a}{a+2(b+c)}$	Tidak ada pasangan 0 - 0 sebagai pembilang atau penyebut. Bobot 2 untuk pasangan
8	$\frac{a}{b+c}$	Perbandingan pasangan yang tidak cocok dengan pasangan 0 - 0

Tabel 2.4

Sebagai contoh, diberikan enam pernyataan kepada pasangan suami istri, sebagai berikut:

- 1) Wanita dengan anak – anak yang masih kecil hanya akan bekerja bila memerlukan uang
- 2) Wanita yang sudah menikah tetap bekerja hanya jika tidak ada masalah dengan pekerjaan rumahnya.
- 3) Wanita yang bekerja tidak boleh meningkatkan prestasi kerja.
- 4) Wanita secara mental dan emosional tidak cocok untuk beberapa pekerjaan.
- 5) Kalau Istri bekerja, suami harus mengerjakan bagian pekerjaan rumah tangga.
- 6) Suami seharusnya menjadi kepala keluarga.

Setiap Pasangan suami istri akan menjawab setiap pernyataan. Jika setuju (1), jika tidak setuju (0). Jawaban responden dicatat dalam matriks respon dan tabel kontingensi. Matriks Respon adalah matriks yang berisi jawaban responden terhadap setiap pernyataan yang diajukan. Tabel Kontingensi merupakan cara untuk menyajikan data kualitatif (non metrik) dengan menyatakan kategori dari

variabel yang satu dalam kategori – kategori dari variabel yang lain. Matriks Respons diberikan pada tabel berikut:

Orang	Sifat					
	1	2	3	4	5	6
Suami	1	0	0	1	1	1
Istri	0	0	0	1	1	0

Tabel 2.5

Dari matrik respons, diperoleh tabel kontingensi sebagai berikut :

	Suami	
Istri	1	0
1	2	0
0	2	2

Tabel Kontingensi

Frekuensi jawaban suami setuju (1) dan istri setuju (1) adalah 2. Frekuensi jawaban suami setuju (1) dan istri tidak setuju (0) adalah 2. Frekuensi jawaban suami tidak setuju (0) dan istri setuju (1) adalah 0. Frekuensi jawaban suami tidak setuju (0) dan istri tidak setuju (0) adalah 2.

Koefisien Asosiasi ditunjukkan sebagai berikut:

$$1) \frac{a+d}{a+b+c+d} = \frac{4}{6} = \frac{2}{3}$$

$$2) \frac{2(a+d)}{2(a+d)+b+c} = \frac{8}{8+2} = \frac{4}{5}$$

$$3) \frac{a+d}{a+d+2(b+c)} = \frac{4}{8} = \frac{1}{2}$$

$$4) \frac{a}{a+b+c+d} = \frac{2}{6} = \frac{1}{3}$$

$$5) \frac{a}{a+b+c} = \frac{2}{4} = \frac{1}{2}$$

$$6) \frac{2a}{2a+b+c} = \frac{4}{6} = \frac{2}{3}$$

$$7) \frac{a}{a+2(b+c)} = \frac{2}{6} = \frac{1}{3}$$

$$8) \frac{a}{b+c} = \frac{2}{2} = 1$$

Koefisien asosiasi yang diperoleh dari perhitungan diatas bervariasi dari 0,33 sampai dengan 1.

Penetapan variabel biner sangat penting dalam menghitung koefisien asosiasi. Variabel biner mempunyai dua kemungkinan hasil. Untuk itu, ketika mengelompokkan orang, variabel yang digunakan: jenis kelamin (Pria / Wanita), kebiasaan merokok (merokok / tidak merokok) dan sebagainya. Dalam matriks data, hasil pengukuran variabel – variabel itu dinyatakan dengan nol (0) atau satu (1). Sebagai contoh, diberikan dua tabel kontingensi dan akan dibandingkan koefisien asosiasi yang diperoleh dari kedua tabel kontingensi.

Tabel kontingensi 1

	1	0
1	1	2
0	1	2

Tabel kontingensi 2

	1	0
1	2	2
0	2	1

Perbandingan koefisien asosiasi dari kedua tabel kontingensi adalah sebagai berikut:

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Tabel 1	1/2	2/3	1/3	1/6	1/4	2/5	1/7	1/3
Tabel 2	1/2	2/3	1/3	1/3	2/5	4/7	1/4	2/3

Tabel 2.6

Dari tabel kontingensi 1 dan tabel kontingensi 2 diperoleh koefisien asosiasi yang berbeda karena variabel biner yang digunakan berbeda.

2.3.3. Koefisien Korelasi

Sebelum membahas koefisien korelasi, terlebih dahulu akan dibahas mengenai nilai harapan, variansi dan kovariansi.

Definisi 2.3.1.

Nilai harapan atau nilai tengah variabel acak X didefinisikan sebagai:

$$E[X] = \sum_x x f_x(x) \text{ jika } X \text{ adalah variabel acak diskret}$$

$$E[X] = \int_R x f_x(x) dx \text{ jika } X \text{ adalah variabel acak kontinu.}$$

dimana $f_x(x)$ adalah fungsi probabilitas variabel acak X . Dalam skripsi ini hanya akan dibahas nilai harapan dengan variabel acak diskret.

Contoh 2.1

Hitung nilai harapan variabel acak Y dari tabel distribusi probabilitas berikut:

Y	$f(y)$
0	1/8
1	1/4
2	3/8
3	1/4

Jawab:

$$E(Y) = \sum_{y=0}^3 yf(y) = (0)(1/8) + (1)(1/4) + (2)(3/8) + (3)(1/4) = 1.75$$

Contoh 2.2

Tentukan nilai harapan bagi X , bila X menyatakan hasil yang diperoleh bila sebuah dadu setimbang dilemparkan.

Jawab :

Masing – masing bilangan 1, 2, 3, 4, 5 dan 6 dapat terjadi dengan peluang 1/6 maka $\mu = E(X) = (1) (1/6) + (2) (1/6) + \dots + (6) (1/6) = 3,5$ artinya, secara rata – rata diperoleh 3,5 per lemparan.



Definisi 2.3.2.

Jika Y_1 dan Y_2 adalah variabel acak, maka kovariansi Y_1 dan Y_2 didefinisikan sebagai nilai harapan dari $(Y_1 - \mu_1)(Y_2 - \mu_2)$ yaitu:

$$Cov(Y_1, Y_2) = E[(Y_1 - \mu_1)(Y_2 - \mu_2)] \text{ dimana } \mu_1 = E(Y_1) \text{ dan } \mu_2 = E(Y_2)$$

Teorema 2.3.1.

Jika X dan Y adalah variabel acak dengan a dan b adalah konstanta maka:

1. $Var(a + bX) = b^2 Var(X)$
2. $Var(aX + bY) = a^2 Var(X) + b^2 Var(Y) + 2ab Cov(X, Y)$

Bukti:

1. Pertama - tama dicari nilai harapan dari $a + bX$ yaitu:

$$\begin{aligned} E(a + bX) &= \sum_x (a + bx)f(x) \\ &= \sum_x af(x) + \sum_x bxf(x) \\ &= a \sum_x f(x) + b \sum_{xx} xf(x) \\ &= a + bE(X) \end{aligned}$$

dapat dicari Variansi dari $a + bX$ yaitu:

$$\begin{aligned} Var(a + bX) &= E\{[a + bX - E(a + bX)]^2\} \\ &= E\{(a + bX - a - b\mu)^2\} \text{ dengan } E(X) = \mu \\ &= E\{b(X - \mu)\}^2 \\ &= E\{b^2(X - \mu)^2\} \end{aligned}$$

$$= b^2 E[(X - \mu)]^2$$

$$= b^2 \text{Var}(X)$$

2. Pertama – tama dicari nilai harapan dari $aX + bY$, yaitu:

$$E(aX + bY) = \sum_x \sum_y (ax + by)f(x, y)$$

$$= \sum_x \sum_y axf(x, y) + \sum_x \sum_y byf(x, y)$$

$$= a \sum_x x \sum_y f(x, y) + b \sum_x y \sum_y f(x, y)$$

$$= a \sum_x xh(x) + b \sum_y yk(y)$$

$$= aE(X) + bE(Y)$$

Sehingga dapat dicari variansi dari $aX + bY$ sebagai berikut:

$$\text{Var}(aX + bY) = E\{[aX + bY - E(aX + bY)]^2\}$$

$$= E\{[(aX - aE(X)) + (bY - bE(Y))]\}^2$$

$$= E\{[a(X - E(X))] + [b(Y - E(Y))]\}^2$$

$$= E\{a^2[X - E(X)]^2 + [b^2[Y - E(Y)]^2 + 2ab[X - E(X)][Y - E(Y)]\}$$

$$= a^2 E[X - E(X)]^2 + b^2 E[Y - E(Y)]^2 + 2abE[X - E(X)][Y - E(Y)]$$

$$= a^2 \text{Var}(X) + b^2 \text{Var}(Y) + 2ab \text{Cov}(X, Y)$$

Teorema 2.3.2.

Misalkan Y_1, \dots, Y_n dan X_1, \dots, X_m adalah variabel acak dengan $E(Y_i) = \mu_i$ dan

$E(X_j) = \xi_j$. Didefinisikan: $U_1 = \sum_{i=1}^n a_i Y_i$ dan $U_2 = \sum_{j=1}^m b_j X_j$ untuk konstanta

$$a_1, \dots, a_n, b_1, \dots, b_m \text{ maka: } Cov(U_1, U_2) = \sum_{i=1}^n \sum_{j=1}^m a_i b_j Cov(Y_i, X_j)$$

Bukti:

$$\begin{aligned} Cov(U_1, U_2) &= E\{[U_1 - E(U_1)][U_2 - E(U_2)]\} \\ &= E\left[\left(\sum_{i=1}^n a_i Y_i - \sum_{i=1}^n a_i \mu_i\right)\left(\sum_{j=1}^m b_j X_j - \sum_{j=1}^m b_j \xi_j\right)\right] \\ &= E\left\{\left[\sum_{i=1}^n a_i (Y_i - \mu_i)\right]\left[\sum_{j=1}^m b_j (X_j - \xi_j)\right]\right\} \\ &= E\left[\sum_{i=1}^n \sum_{j=1}^m a_i b_j (Y_i - \mu_i)(X_j - \xi_j)\right] \\ &= \sum_{i=1}^n \sum_{j=1}^m a_i b_j E(Y_i - \mu_i)(X_j - \xi_j) \\ &= \sum_{i=1}^n \sum_{j=1}^m a_i b_j Cov(Y_i, X_j) \end{aligned}$$

Koefisien korelasi merupakan salah satu alat untuk mengukur similaritas antar obyek. Misalkan ada dua variabel acak X dan Y dengan X adalah variabel umur suatu mobil bekas dan Y adalah variabel nilai jual mobil tersebut, maka untuk nilai X yang besar berpadanan dengan nilai Y yang kecil dan sebaliknya untuk nilai X yang kecil berpadanan dengan nilai Y yang besar.

Definisi 2.3.3.

Koefisien korelasi antara dua variabel acak didefinisikan sebagai berikut:

$$\rho_{xy} = \frac{Cov[X, Y]}{\sqrt{Var[X]}\sqrt{Var[Y]}}$$

Teorema 2.3.3.

Koefisien Korelasi ρ_{XY} bernilai diantara -1 dan 1 yaitu $|\rho_{XY}| \leq 1$.

Bukti:

Jika (X, Y) menyatakan variabel acak bivariat dengan $E[X] = \mu_1$, $E[Y] = \mu_2$,

$Var[X] = \sigma_1^2$, $Var[Y] = \sigma_2^2$, dan $Cov[X, Y] = \sigma_{12}$.

Didefinisikan dua variabel acak baru: $U = \frac{X - \mu_1}{\sigma_1}$ dan $V = \frac{Y - \mu_2}{\sigma_2}$ maka $E[U] =$

$E[V] = 0$, $E[U^2] = E[V^2] = 1$ dan

$$\begin{aligned} \rho_{xy} &= \frac{\sigma_{12}}{\sigma_1 \sigma_2} = \frac{E[(X - \mu_1)(Y - \mu_2)]}{\sigma_1 \sigma_2} \\ &= E\left[\frac{(X - \mu_1)}{\sigma_1} \frac{(Y - \mu_2)}{\sigma_2}\right] = E[UV] \end{aligned}$$

Kemudian dipertimbangkan $E[(U - V)^2]$ Nilai harapan ini positif dan tepat sama dengan nol jika $\Pr[U = V] = 1$. Oleh karena itu,

$$\begin{aligned} 0 &\leq E[(U - V)^2] = E[U^2 + V^2 - 2UV] \\ &= E[U^2] + E[V^2] - 2E[UV] = 1 + 1 - 2\rho_{xy} \\ &= 2 - 2\rho_{xy} \end{aligned}$$

yang berimplikasi $\rho_{xy} \leq 1$.

Analog dengan bukti diatas, $E[(U + V)^2] \geq 0$

$$E[U^2 + V^2 + 2UV] \geq 0$$

$$E(U^2) + 2E(UV) + E(V^2) \geq 0$$

$$1 + 2\rho_{xy} + 1 \geq 0$$

$$2 + 2\rho_{xy} \geq 0 \text{ mempunyai implikasi } \rho_{xy} \geq -1$$

Jadi, terbukti $-1 \leq \rho_{xy} \leq 1$

Dicoba dengan persamaan $Y = a + bX$ untuk konstan a dan b , dengan $b \neq 0$

Untuk menunjukkan $\rho_{xy} = 1$ dinyatakan dengan :

$$\frac{X - \mu_1}{\sigma_1} = \frac{Y - \mu_2}{\sigma_2}$$

$$Y - \mu_2 = \frac{\sigma_2}{\sigma_1}(X - \mu_1)$$

$$Y = \mu_2 + \frac{\sigma_2}{\sigma_1}(X - \mu_1)$$

$$Y = \mu_2 + \frac{\sigma_2}{\sigma_1}X - \frac{\sigma_2}{\sigma_1}\mu_1$$

$$Y = \left(\mu_2 - \frac{\sigma_2}{\sigma_1}\mu_1 \right) + \frac{\sigma_2}{\sigma_1}X$$

dari persamaan $Y = a + bX$ diperoleh, $a = \mu_2 - \frac{\sigma_2}{\sigma_1}\mu_1$ dan $b = \frac{\sigma_2}{\sigma_1}$ artinya

persamaan $Y = a + bX$ mempunyai kemiringan positif (korelasi positif).

Untuk menunjukkan $\rho_{xy} = -1$ dinyatakan dengan :

$$\frac{X - \mu_1}{\sigma_1} = -\frac{Y - \mu_2}{\sigma_2}$$

$$-Y + \mu_2 = \frac{\sigma_2}{\sigma_1}(X - \mu_1)$$

$$Y = \mu_2 - \frac{\sigma_2}{\sigma_1}(X - \mu_1)$$

$$Y = \mu_2 - \frac{\sigma_2}{\sigma_1} X + \frac{\sigma_2}{\sigma_1} \mu_1$$

$$Y = \left(\mu_2 + \frac{\sigma_2}{\sigma_1} \mu_1 \right) - \frac{\sigma_2}{\sigma_1} X$$

dari persamaan $Y = a + bX$ diperoleh, $a = \mu_2 + \frac{\sigma_2}{\sigma_1} \mu_1$ dan $b = -\frac{\sigma_2}{\sigma_1}$ artinya

persamaan $Y = a + bX$ mempunyai kemiringan negatif (korelasi negatif)

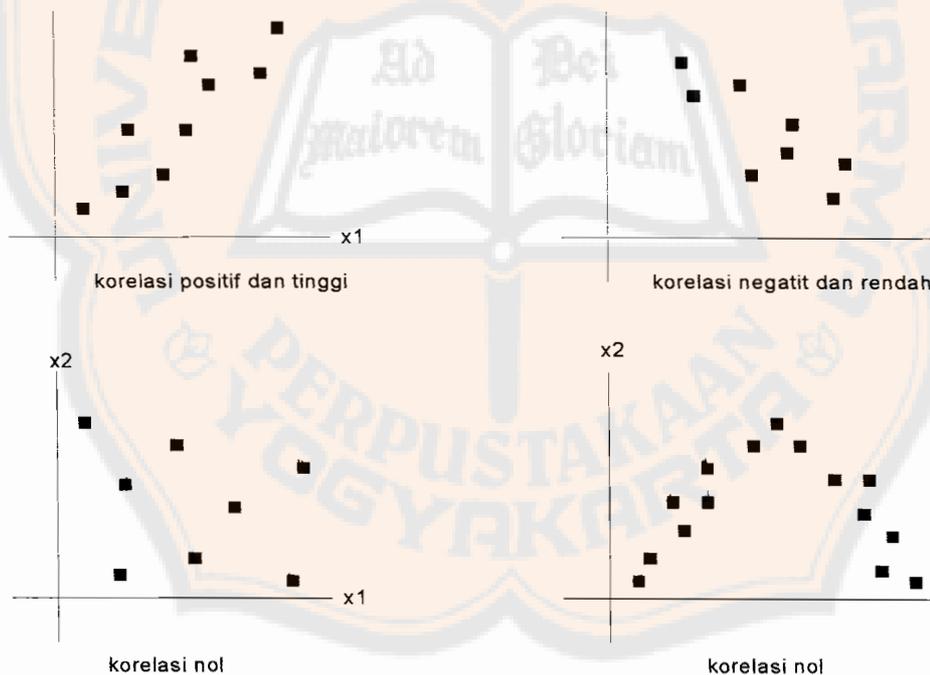
Jadi dengan persamaan $Y = a + bX$ dapat dibuktikan $|\rho_{xy}| = 1$

Tingkat keeratan hubungan antara variabel X dan Y diukur dengan koefisien korelasi. Koefisien korelasi linear didefinisikan sebagai ukuran hubungan linear antara dua variabel acak X dan Y dan dilambangkan dengan ρ_{xy} . Jadi, ρ_{xy} mengukur sejauh mana titik – titik menggerombol sekitar sebuah garis lurus. Oleh karena itu, dengan membuat diagram pencar bagi n pengamatan $\{(X_i, Y_i); i = 1, 2, \dots, n\}$ dalam sampel acak, dapat ditarik kesimpulan tertentu mengenai ρ_{xy} . Bila titik – titik menggerombol mengikuti sebuah garis lurus dengan kemiringan positif, maka ada korelasi positif yang tinggi antara kedua variabel. Akan tetapi bila titik – titik menggerombol mengikuti sebuah garis lurus dengan kemiringan negatif, maka antara kedua variabel itu terdapat korelasi negatif yang tinggi. Korelasi antara kedua variabel semakin menurun secara numerik dengan semakin memencarnya atau menjauhnya titik – titik dari suatu garis lurus. Bila titik – titik mengikuti suatu pola yang acak, dengan kata lain

tidak ada pola, maka kedua variabel tersebut mempunyai korelasi nol, dan disimpulkan tidak ada hubungan linear antara X dan Y .

Karena koefisien korelasi antara dua variabel adalah suatu tingkat keeratan hubungan linear antara kedua variabel tersebut, sehingga bila $\rho_{xy} = 0$ maka tidak ada hubungan linear, bukan berarti bahwa antar kedua variabel tersebut pasti tidak terdapat hubungan. Jadi bila antara X dan Y terdapat suatu hubungan kuadratik maka diperoleh korelasi nol meskipun jelas ada hubungan tidak linear antara kedua variabel tersebut.

Ukuran korelasi linear antara dua variabel yang paling banyak digunakan adalah koefisien korelasi momen - hasil kali Pearson atau koefisien korelasi sampel.



Gambar 2.3

Contoh 2.3

Diberikan Fungsi probabilitas bivariat variabel acak X dan Y adalah sebagai berikut:

	X	0	1
y			
2		0.4	0.1
1		0.1	0.4

Hitung dan tafsirkan koefisien korelasi bagi data yang diberikan:

Jawab:

$$\mu_x = (0)(0.5) + (1)(0.5) = 0.5$$

$$\mu_y = (2)(0.5) + (1)(0.5) = 1.5$$

$$Var(X) = E[X^2] - (E[X])^2 = (0^2)(0.5) + (1^2)(0.5) - (0.5)^2 = 0.25$$

$$Var(Y) = E[Y^2] - (E[Y])^2 = (1^2)(0.5) + (2^2)(0.5) - (1.5)^2 = 0.25$$

$$E[XY] = (0)(2)(0.4) + (0)(1)(0.1) + (1)(2)(0.1) + (1)(1)(0.4) = (0.6)$$

$$\text{maka Cov}(X, Y) = E[XY] - E[X]E[Y] = 0.6 - (0.5)(1.5) = -0.15$$

$$\text{Jadi, koefisien korelasi } \rho_{xy} = \frac{-0.15}{(0.5)(0.5)} = \frac{-0.15}{(0.25)} = -0.60$$

2.4 Analisis Variansi (ANOVA)

Analisis variansi (ANOVA) adalah suatu teknik analisis data untuk membandingkan rata – rata populasi normal dimana hanya ada satu variabel berskala metrik tak bebas yang akan di analisis.

Langkah – langkah dalam analisis vvariansi adalah sebagai berikut:

1. Menyatakan hipotesis nol $H_0 : \mu_1 = \mu_2 = \dots = \mu_k$

2. Menentukan hipotesis alternatif $H_1 : \exists \mu_i \neq \mu_j, i \neq j$
3. Menentukan taraf signifikansi α
4. Menggunakan statistik uji F kemudian menentukan wilayah kritiknya
 $f > f_\alpha[k-1, k(n-1)]$.
5. Menghitung nilai statistik uji berdasarkan data sampel.
6. Keputusan: Tolak H_0 bila nilai statisti uji tersebut jatuh dalam wilayah kritiknya.

Andaikan dari k populasi normal diambil sampel berukuran n dengan rata – rata $\mu_1, \mu_2, \dots, \mu_k$ dan variansi σ^2 . x_{ij} adalah pengamatan ke j dari populasi ke i .

Data diberikan dalam tabel berikut:

	Populasi						
	1	2	...	i	...	k	
	x_{11}	x_{21}	...	x_{i1}	...	x_{k1}	
	x_{12}	x_{22}	...	x_{i2}	...	x_{k2}	
	\vdots	\vdots		\vdots		\vdots	
	x_{1n}	x_{2n}	...	x_{in}	...	x_{kn}	
Total	$T_{1.}$	$T_{2.}$...	$T_{i.}$...	$T_{k.}$	$T_{..}$
Rata - rata	$\bar{x}_{1.}$	$\bar{x}_{2.}$...	$\bar{x}_{i.}$...	$\bar{x}_{k.}$	$\bar{x}_{..}$

Tabel 2.7

Dimana:

$T_{i.}$ adalah total pengamatan dalam sampel dari populasi ke i .

$\bar{x}_{i.}$ adalah rata – rata semua pengamatan dalam sample dari popullasi ke i .

$T_{..}$ adalah total semua nk pengamatan.

$\bar{x}_{..}$ adalah rata – rata semua nk pengamatan.

Tabel analisis variansi diberikan sebagai berikut:

Sumber varian	Jumlah Kuadrat	Derajat Bebas	Kuadrat Tengah	f Hitung
Kelompok	JKK	k - 1	$s_1^2 = \frac{JKK}{k-1}$	$\frac{s_1^2}{s_2^2}$
Galat	JKG	k (n - 1)	$s_2^2 = \frac{JKG}{k(n-1)}$	
Total	JKT	nk - 1		

Tabel 2.8

Dengan Rumus Hitung Jumlah Kuadrat:

$$\mathbf{JKT} \text{ (Jumlah Kuadrat Total)} = \sum_{i=1}^k \sum_{j=1}^{n_j} x_{ij}^2 - \frac{T_{..}^2}{N},$$

$$\mathbf{JKK} \text{ (Jumlah Kuadrat rata - rata kolom)} = \frac{\sum_{i=1}^k T_{i.}^2}{n_i} - \frac{T_{..}^2}{N},$$

$$\mathbf{JKG} \text{ (Jumlah Kuadrat Galat)} = \mathbf{JKT} - \mathbf{JKK}$$

Statistik uji yang digunakan adalah distribusi F dengan ketentuan: Tolak H_0 pada taraf signifikan α bila $f > f_{\alpha}[k-1, k(n-1)]$

ANOVA akan digunakan dalam aplikasi analisis kelompok yaitu untuk menginterpretasi kelompok dengan menggunakan program SPSS. Dengan melihat nilai Significance (Sig), suatu variabel dikatakan membedakan kelompok jika mempunyai nilai Sig < 0,05.

BAB III ANALISIS KELOMPOK

3.1. PENDAHULUAN

Dalam suatu penelitian, kadang peneliti perlu mengelompokkan obyek-obyek berdasarkan similaritas karakteristik yang dimiliki oleh obyek-obyek tersebut. Obyek-obyek itu dapat berupa produk (barang dan jasa), orang (responden, konsumen). Nama umum teknik pengelompokan obyek kedalam beberapa kelompok berdasarkan karakteristik tertentu disebut analisis kelompok (*Cluster Analysis*).

Obyek-obyek yang berada dalam satu kelompok akan mempunyai similaritas satu dengan yang lain pada karakteristik tersebut. Dengan kata lain, variansi dalam kelompok lebih kecil daripada variansi dengan kelompok yang lain. Jadi suatu kelompok yang terbentuk dikatakan baik jika memiliki ciri-ciri sebagai berikut:

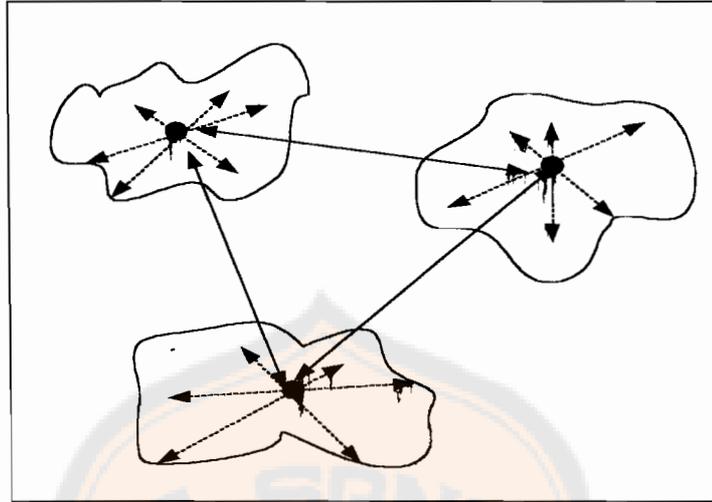
1. Homogenitas (similaritas) yang tinggi antar anggota dalam satu kelompok.
2. Heterogenitas (perbedaan) yang tinggi antar kelompok yang satu dengan kelompok yang lain.

Hal ini, dapat digambarkan dalam diagram berikut:

Keterangan gambar:

↔ : variasi antar kelompok

↔ : variasi dalam satu kelompok



Gambar 3.1

Analisis kelompok dapat diterapkan dalam berbagai bidang ilmu, antara lain:

- Psikologi
Analisis kelompok membantu melakukan pengelompokan orang berdasarkan respon mereka terhadap stimulus tertentu yang diberikan atau pengelompokan orang berdasarkan kepribadian yang dimiliki.
- Biologi
Analisis kelompok membantu proses taxonomi untuk mengelompokkan organisme tertentu.
- Manajemen
Analisis kelompok membantu mengelompokkan konsumen produk tertentu berdasarkan pendapatan yang diperoleh.
- Pendidikan
Analisis kelompok membantu mengelompokkan siswa yang gagal atau berhasil di sekolah berdasarkan hasil tes yang diperoleh.

Sebenarnya, masih banyak lagi penerapan analisis kelompok ini. Dalam skripsi ini, penulis hanya akan membahas penerapan analisis kelompok dalam bidang Pendidikan.

Akan diberikan definisi mengenai analisis kelompok sebagai berikut:

Definisi 3.1

Analisis Kelompok adalah teknik pengelompokan multivariat yang digunakan untuk mengelompokkan obyek berdasarkan karakteristik yang dimiliki.

Dalam analisis kelompok, konsep variat merupakan persoalan utama. Kelompok variat (*Cluster Variate*) adalah himpunan variabel yang mewakili karakteristik yang digunakan untuk membandingkan obyek, maka kelompok variat menentukan 'karakter' obyek tersebut.

Analisis Kelompok adalah teknik multivariat yang tidak hanya menaksir variat secara empiris, tetapi lebih daripada itu peneliti juga dapat menggunakan variat yang telah ditetapkan sendiri.

Titik utama analisis kelompok adalah persamaan obyek berdasar pada variat bukan pada taksiran variat itu sendiri. Hal ini menjadikan pendefinisian variat oleh peneliti sebagai suatu langkah yang penting dalam analisis kelompok.

3.2. PROSES DASAR ANALISIS KELOMPOK

Adapun proses yang harus dilakukan dalam analisis kelompok adalah:

1. Mengukur Similaritas Antar Obyek.
2. Proses Standarisasi
3. Memilih Algoritma Pengelompokan

4. Interpretasi Kelompok
5. Membuat Profil Kelompok

3.2.1. Mengukur Similaritas Antar Obyek

Analisis kelompok pada dasarnya mengelompokkan obyek yang mempunyai similaritas. Dalam proses ini, diukur seberapa jauh similaritas yang dimiliki antar obyek. Ada tiga ukuran derajat similaritas antar obyek berdasarkan variabel tertentu.

1. Korelasi antar sepasang obyek pada beberapa variabel.
2. Jarak antar dua obyek. Ada beberapa macam ukuran jarak, antara lain Jarak Euclidean, Jarak Mahalanobis, Jarak Minkowski.
3. Koefisien Assosiasi antar obyek (*Association Measures*)

Dari ketiga ukuran derajat similaritas diatas, korelasi dan jarak digunakan jika jenis datanya metrik. Sedangkan Koefisien Assosiasi digunakan jika jenis datanya nonmetrik.

3.2.2. Proses Standarisasi.

Proses Standarisasi adalah proses yang mengubah setiap skor data mentah ke dalam nilai standar yang mempunyai rata-rata = 0 dan standar deviasi = 1. Bentuk standarisasi adalah transformasi setiap variabel kedalam skor standar (Z skor). Standarisasi merupakan bentuk umum dari fungsi jarak normal, yang menggunakan ukuran Jarak Euclidean untuk transformasi normal data mentah.

Setelah ukuran similaritas dipilih timbul suatu pertanyaan “Haruskah melakukan proses standarisasi untuk semua data sebelum menghitung similaritasnya?” Dalam menjawab pertanyaan ini, peneliti harus mempertimbangkan “Apakah ukuran jarak mempunyai perbedaan skala / perbedaan jarak antar variabel / perbedaan obyek yang sangat sensitif atau tidak” Pada umumnya, data yang variansinya besar (standar deviasi besar) mempunyai pengaruh yang lebih besar pada nilai similaritas akhir. Sebagai contoh, individu – individu dikelompokan berdasar tiga variabel yaitu sikap terhadap produk tertentu (suka / tidak suka), umur (tahun), dan pendapatan (dolar). Jarak antar obyek (similaritas) hamper semuanya sama dengan selisih pendapatan.

Bentuk standarisasi kedalam skor standar (Z skor) yaitu:

$$Z = \frac{x_i - \mu}{\sigma} \quad \text{atau} \quad Z = \frac{x_i - \bar{X}}{S}$$

Dimana:

x_i	= nilai / skor data mentah	Z	= nilai standar
μ	= rata-rata populasi	\bar{X}	= rata – rata sampel
σ	= Standar deviasi populasi	S	= standar deviasi sampel

Ada beberapa keuntungan yang diperoleh dari proses standarisasi data, antara lain:

1. Mudah membandingkan variabel – variabel pada skala yang sama (nilai rata-rata = 0 dan nilai standar deviasi = 1)

2. Tidak ada perbedaan nilai standarisasi jika hanya merubah skalanya. Misal waktu (dalam menit) dengan proses standarisasi diubah dalam detik nilai standarisasinya adalah sama.

Akan tetapi peneliti tidak harus selalu memakai standarisasi data tanpa mempertimbangkan konsekuensinya. Proses standarisasi bukan alasan mutlak, kecuali penyelesaian pengelompokan harus menggunakan variabel yang distandarisasi. Jadi keputusan untuk melakukan proses standarisasi harus dibuat dengan pertimbangan yang hati-hati. Sebagai contoh, diberikan data sebagai berikut:

Obyek	Sikap Terhadap Produk		
	Tertentu	Umur (Th)	Pendapatan (Rp)
A	1	26	450000
B	0	35	700000
C	1	42	850000
D	1	40	400000

Tabel 3.1

Pada variabel sikap, (1) menyatakan suka dan (0) menyatakan tidak suka.

Secara umum hasil pengukuran tersebut dapat dinyatakan dalam matriks $n \times p$ sebagai berikut:

$$\begin{matrix}
 & & & & p \text{ variabel} \\
 n \text{ obyek} & \begin{bmatrix}
 x_{11} & \cdots & x_{1f} & \cdots & x_{1p} \\
 \vdots & & \vdots & & \vdots \\
 x_{i1} & \cdots & x_{if} & \cdots & x_{ip} \\
 \vdots & & \vdots & & \vdots \\
 x_{n1} & \cdots & x_{nf} & \cdots & x_{np}
 \end{bmatrix}
 \end{matrix}$$

dengan $i = 1, \dots, n$

$$f = 1, \dots, p$$

Dari data tersebut, jarak dihitung dengan menggunakan Jarak Euclidean. Untuk membandingkan jarak, dilakukan dengan menggunakan dua cara: data distandarisasi atau data tanpa distandarisasi.

- Dengan Proses Standarisasi.

Perhitungan: $\bar{X}_f = \frac{1}{n}(x_{1f} + x_{2f} + \dots + x_{nf})$

$$\bar{X}_1 = \frac{1}{4}(1+0+1+1) = \frac{3}{4} = 0,75$$

Analog dengan \bar{X}_1 diperoleh $\bar{X}_2 = 35,75$ dan $\bar{X}_3 = 600000$

$$S_f = \sqrt{\frac{1}{n-1} \{ (x_{1f} - \bar{X}_f)^2 + (x_{2f} - \bar{X}_f)^2 + \dots + (x_{nf} - \bar{X}_f)^2 \}}$$

$$S_1 = \sqrt{\frac{1}{3} \{ (1-0,75)^2 + (0-0,75)^2 + (1-0,75)^2 + (1-0,75)^2 \}}$$

$$= \sqrt{\frac{1}{3} (0,0625 + 0,5625 + 0,0625 + 0,0625)}$$

$$= \sqrt{\frac{1}{3} (0,75)}$$

$$= \sqrt{0,25}$$

$$= 0,5$$

Analog dengan perhitungan S_1 , diperoleh $S_2 = 7,136$ dan $S_3 = 212132,03$

Nilai standar diperoleh dengan menggunakan rumus berikut:

$$Z_{if} = \frac{x_{if} - \overline{X}_f}{S_f}$$

$$Z_{11} = \frac{1 - 0,75}{0,5} = \frac{0,25}{0,5} = 0,5$$

Analog dengan perhitungan Z_{11} , diperoleh nilai standar yang lain adalah sebagai berikut:

$$Z_{21} = -1,5 \quad Z_{31} = 0,5 \quad Z_{41} = 0,5$$

$$Z_{12} = -1,366 \quad Z_{22} = -0,105 \quad Z_{32} = 0,876 \quad Z_{42} = 0,595$$

$$Z_{13} = -0,707 \quad Z_{23} = 0,471 \quad Z_{33} = 1,178 \quad Z_{43} = -0,943$$

Nilai standar yang diperoleh dimasukkan kedalam matriks berikut:

Obyek	Sikap Terhadap Produk Tertentu	Umur (Th)	Pendapatan (Rp)
A	0.5	-1.366	-0.707
B	-1.5	-0.105	0.471
C	0.5	0.876	1.178
D	0.5	0.595	-0.946

Tabel 3.2

Selanjutnya, jarak dihitung dengan menggunakan Jarak Euclidean sebagai

berikut:
$$d(i, j) = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \dots + (x_{ip} - x_{jp})^2}$$

$$\begin{aligned} d(AB) &= \sqrt{(0.5 + 1.5)^2 + (-1.366 + 0.105)^2 + (-0.707 - 0.471)^2} \\ &= \sqrt{4 + 1.59 + 1.388} \\ &= \sqrt{6.977684} \\ &= 2.64 \end{aligned}$$

Analog dengan jarak antara obyek A dengan B diperoleh:

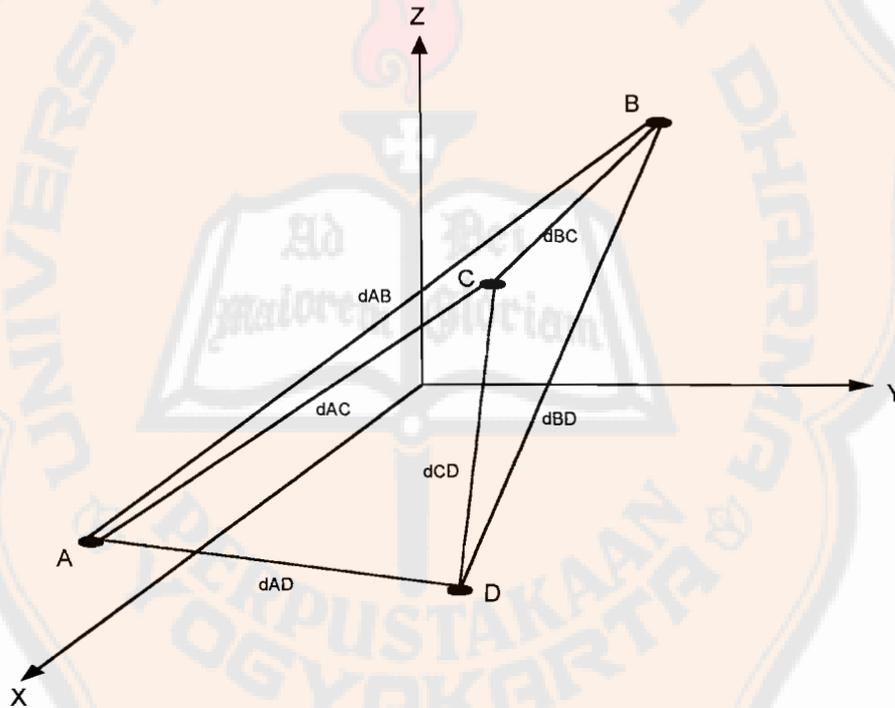
$$d(AC) = 2,929 \quad d(BD) = 2,547 \quad d(AD) = 1,975$$

$$d(CD) = 2,139 \quad d(BC) = 2,337$$

Besarnya jarak yang diperoleh dimasukkan kedalam matriks jarak berikut:

	A	B	C	D
A	0			
B	2,64	0		
C	2,929	2,337	0	
D	1,975	2,547	2,139	0

Hasil yang diperoleh, kemudian digambarkan pada ruang dimensi tiga berikut:



Gambar 3.2

- Tanpa Proses Standarisasi

$$\begin{aligned}
 d(AB) &= \sqrt{(1-0)^2 + (26-35)^2 + (450000-700000)^2} \\
 &= \sqrt{1+81+625000000} \\
 &= \sqrt{6250000082} \\
 &= 250000,0002
 \end{aligned}$$

Analog dengan jarak antara obyek *A* dan obyek *B*, diperoleh

$$\begin{aligned}
 d(AC) &= 400000,0003 & d(BD) &= 300000 & d(AD) &= \\
 & & & & & 50000,00196
 \end{aligned}$$

$$\begin{aligned}
 d(CD) &= 450000 & d(BC) &= 150000,0002
 \end{aligned}$$

Besarnya jarak yang diperoleh dimasukkan kedalam matriks jarak berikut:

	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>
<i>A</i>	0			
<i>B</i>	250000,0002	0		
<i>C</i>	400000,0003	150000,0002	0	
<i>D</i>	50000,00196	300000	450000	0

Jarak dari hasil perhitungan untuk data yang tidak melalui proses standarisasi, hampir semua nilai jarak antar 2 obyek sama dengan selisih pendapatan (Rp). Misalnya, jarak antara obyek *C* dan obyek *D* yaitu $d(CD) = 450.000$ sama dengan pendapatan obyek *C* – pendapatan obyek *D* = 850000 – 400000.

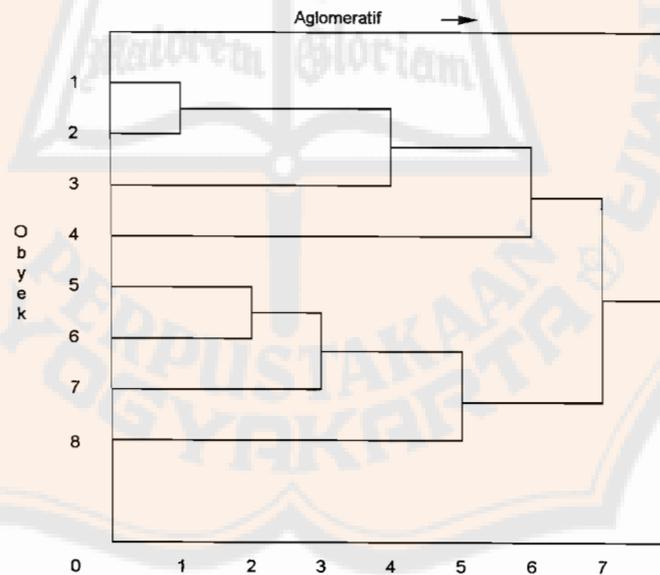
3.2.3. Memilih Algoritma Pengelompokan

Ada dua metode yang digunakan dalam proses ini, yaitu:

3.2.3.1. Metode Hierarki (Bertingkat)

Dalam metode ini, pengelompokan dimulai dengan mengukur dua atau lebih obyek yang memiliki derajat similaritas tertinggi / paling dekat, kemudian dilanjutkan dengan obyek lain yang mempunyai derajat similaritas tertinggi kedua (kedekatan kedua).

Proses ini dilakukan terus menerus hingga terbentuk semacam diagram (dendogram) ‘pohon’ sehingga terlihat dengan jelas adanya tingkatan antar obyek dari yang paling mirip sampai yang paling tidak mirip. Pada akhirnya, semua obyek akan membentuk sebuah kelompok besar. Metode pengelompokan seperti ini disebut metode Aglomeratif. Untuk membantu memperjelas proses Hierarki ini diberikan ilustrasi dendogram berikut:



Dendogram 3.1

Ada lima metode populer yang digunakan dalam metode Aglomeratif untuk membuat kelompok. Kelima metode itu adalah:

3.2.3.1.1. Single Linkage

Metode Single Linkage akan mengelompokkan dua obyek yang mempunyai jarak terdekat terlebih dahulu. Kemudian, jarak terdekat yang kedua diperoleh dari obyek ketiga dengan dua obyek yang telah menjadi satu kelompok untuk membentuk kelompok baru.

Jika C_1 dan C_2 adalah dua kelompok, maka jarak antara C_1 dan C_2 didefinisikan sebagai dissimilaritas terkecil antara anggota C_1 dan anggota C_2 dan dinyatakan dengan:

$$d_{(C_1)(C_2)} = \min \{d_{rs} : r \in C_1, s \in C_2\}$$

Contoh 3.1

Untuk menggambarkan algoritma Single Linkage, diberikan matriks jarak 5 pasang obyek sebagai berikut:

$$D = \{d_{ik}\} = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{matrix} & \begin{bmatrix} 0 & & & & \\ 9 & 0 & & & \\ 3 & 7 & 0 & & \\ 6 & 5 & 9 & 0 & \\ 11 & 10 & 2 & 8 & 0 \end{bmatrix} \end{matrix}$$

Andaikan setiap obyek diperlakukan sebagai kelompok. Pengelompokan dimulai dengan menggabungkan dua obyek terdekat, dimana $(d_{ik})_{\min} = d_{53} = 2$. Obyek 5 dan 3 bergabung membentuk kelompok (35). Untuk pengelompokan

selanjutnya diperlukan jarak antara kelompok (35) dengan obyek yang lain 1, 2

dan 4. Dihitung jaraknya sebagai berikut:

$$d_{(35)1} = \min\{d_{31}, d_{51}\} = \min\{3, 1\} = 3$$

$$d_{(35)2} = \min\{d_{32}, d_{52}\} = \min\{3, 10\} = 7$$

$$d_{(35)4} = \min\{d_{34}, d_{54}\} = \min\{9, 8\} = 8$$

Penghapusan baris dan kolom matriks D yang menghubungkan obyek 3 dan 5 serta penambahan baris dan kolom untuk kelompok (35) menghasilkan matriks jarak yang baru sebagai berikut:

$$D = \{d_{ik}\} = \begin{matrix} & \begin{matrix} (35) & 1 & 2 & 3 \end{matrix} \\ \begin{matrix} (35) \\ 1 \\ 2 \\ 4 \end{matrix} & \begin{bmatrix} 0 & & & \\ 3 & 0 & & \\ 7 & 9 & 0 & \\ 8 & 6 & 5 & 0 \end{bmatrix} \end{matrix} \quad \text{Jarak terdekat antar pasangan}$$

kelompok adalah $d_{(35)1} = 3$ dan gabungan kelompok (35) dengan obyek 1 menghasilkan kelompok (135). Jarak kelompok (135) dengan obyek 2 dan 4 adalah:

$$d_{(135)2} = \min\{d_{(35)2}, d_{12}\} = \min\{7, 9\} = 7$$

$$d_{(135)4} = \min\{d_{(35)4}, d_{14}\} = \min\{8, 6\} = 6$$

Matriks jarak untuk tingkat pengelompokan selanjutnya adalah

$$D = \{d_{ik}\} = \begin{matrix} & \begin{matrix} (135) & 2 & 4 \end{matrix} \\ \begin{matrix} (135) \\ 2 \\ 4 \end{matrix} & \begin{bmatrix} 0 & & \\ 7 & 0 & \\ 6 & 5 & 0 \end{bmatrix} \end{matrix}$$

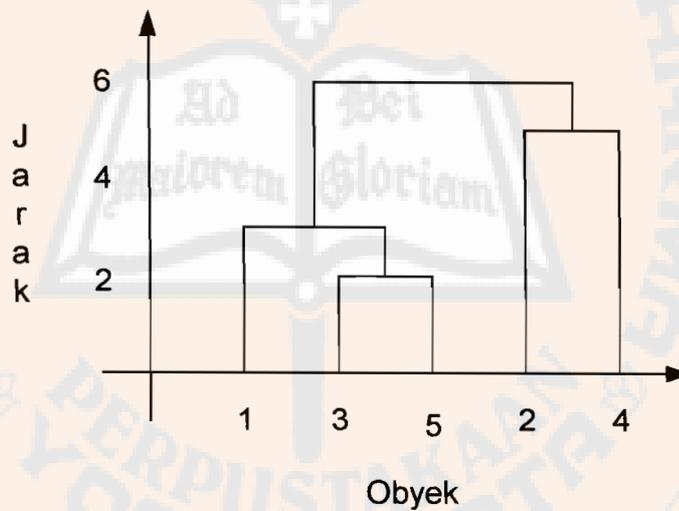
Jarak terdekat antar pasangan kelompok adalah $d_{(42)} = 5$. Gabungan obyek 4 dan 2 membentuk kelompok (24) sehingga terbentuk dua kelompok yang berbeda yaitu (135) dan (24) yang jaraknya,

$$d_{(135)(24)} = \min \{d_{(135)2}, d_{(135)4}\} = \min \{7, 6\} = 6$$

Diperoleh matriks jaraknya sebagai berikut:

$$D = \{d_{(ik)}\} = \begin{matrix} & \begin{matrix} (135) & (24) \end{matrix} \\ \begin{matrix} (135) \\ (24) \end{matrix} & \begin{bmatrix} 0 & \\ 6 & 0 \end{bmatrix} \end{matrix}$$

Oleh karena itu, dengan jarak terdekat = 6 kelompok (135) dan (24) bergabung membentuk kelompok tunggal yang meliputi kelima obyek (12345) Proses pengelompokan tersebut dapat digambarkan pada dendogram berikut:



Dendogram 3.2

3.2.3.1.2 Complete Linkage

Prosedur pengelompokan pada metode Complete Linkage (*Furthest Neighbor*) berbeda dengan metode Single Linkage dalam hal jarak antara dua kelompok yang didefinisikan sebagai dissimilaritas terbesar antara anggota C_1 dan anggota C_2 dan didefinisikan sebagai berikut:

$$d_{(C_1)(C_2)} = \max\{d_{rs} : r \in C_1, s \in C_2\}$$

dimana r menyatakan obyek r dan s menyatakan obyek s . Pada setiap langkah, dua kelompok terdekat digabungkan terlebih dahulu.

Contoh 3.2

Diberikan matriks jarak sebagai berikut:

$$D = \{d_{ik}\} = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{matrix} & \begin{bmatrix} 0 & 9 & 3 & 6 & 11 \\ & 0 & 7 & 5 & 10 \\ & & 0 & 9 & 2 \\ & & & 0 & 8 \\ & & & & 0 \end{bmatrix} \end{matrix}$$

Pada langkah pertama, obyek 3 dan 5 bergabung menjadi kelompok (35).

Selanjutnya, dihitung:

$$d_{(35)1} = \max\{d_{31}, d_{51}\} = \max\{3, 11\} = 11$$

$$d_{(35)2} = \max\{d_{32}, d_{52}\} = \max\{7, 10\} = 10$$

$$d_{(35)4} = \max\{d_{34}, d_{54}\} = \max\{9, 8\} = 9$$



Matriks jarak menjadi:

$$(35) \quad 1 \quad 2 \quad 4$$

$$D = \{d_{ik}\} = \begin{matrix} (35) \\ 1 \\ 2 \\ 4 \end{matrix} \begin{bmatrix} 0 & & & \\ 11 & 0 & & \\ 10 & 9 & 0 & \\ 9 & 6 & 5 & 0 \end{bmatrix}$$

Terbentuk kelompok (24) setelah obyek 2 bergabung dengan obyek 4.

Selanjutnya, dihitung:

$$d_{(24)(35)} = \max\{d_{2(35)}, d_{4(35)}\} = \max\{10, 9\} = 10$$

$$d_{(24)1} = \max\{d_{21}, d_{41}\} = \max\{9, 6\} = 9$$

Matriks jarak menjadi:

$$(35) \quad 1 \quad (24)$$

$$D = \{d_{ik}\} = \begin{matrix} (35) \\ 1 \\ (24) \end{matrix} \begin{bmatrix} 0 & & \\ 11 & 0 & \\ 10 & 9 & 0 \end{bmatrix}$$

Selanjutnya terbentuk kelompok (124) dengan jarak = 9.

Langkah terakhir, kelompok (35) akan bergabung dengan kelompok (124)

membentuk satu kelompok tunggal (12345).

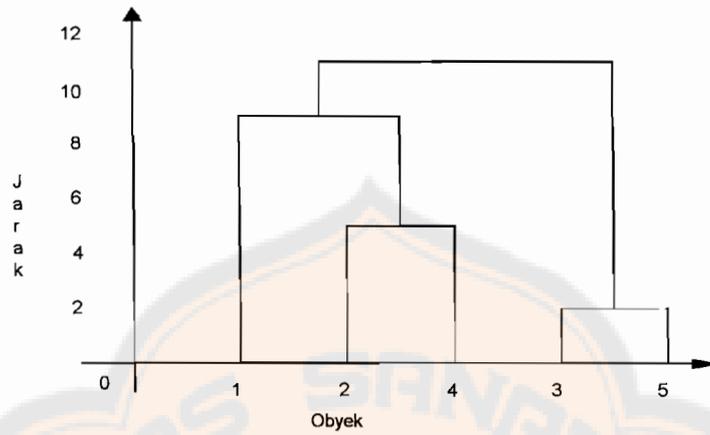
$$d_{(124)(35)} = \max\{d_{1(35)}, d_{(24)(35)}\} = \max\{1, 10\} = 11$$

Sehingga matriks jarak menjadi:

$$(135)(24)$$

$$D = \{d_{(ik)}\} = \begin{matrix} (135) \\ (24) \end{matrix} \begin{bmatrix} 0 & \\ 11 & 0 \end{bmatrix}$$

Dendrogram Complete Linkage untuk jarak antara lima obyek adalah sebagai berikut:



Dendrogram 3.3

3.2.3.1.3 Average Linkage

Metode Average Linkage mengelompokkan obyek berdasar jarak rata – rata yang didapat dengan menentukan rata – rata semua jarak antar obyek terlebih dahulu.

Algoritma Average Linkage dimulai dengan mencari matriks jarak $d_{(C_1)(C_2)}$ untuk menentukan obyek dengan jarak terdekat (similaritas terbesar). Jarak antara C_1 dan C_2 didefinisikan sebagai rata – rata $n_1 n_2$ buah dissimilaritas antara semua pasangan obyek. Didefinisikan sebagai berikut:

$$d_{(C_1)(C_2)} = \frac{1}{n_1 n_2} \sum_{r \in C_1} \sum_{s \in C_2} d_{rs}$$

Contoh 3.3

Diberikan matriks jarak yang menyatakan derajat similaritas antara 11 kelompok bahasa. Nilai jarak berkisar antara 0 sampai 10. Semakin besar nilai jarak, derajat similaritas semakin kecil. Jadi, nilai 0 menyatakan derajat similaritas besar (sama) dan nilai 10 menyatakan derajat similaritas sangat kecil.

Dimana,

- E = Inggris Sp = Spanyol N = Norwegia
- I = Italia Da = Denmark P = Polandia
- Du = Belanda H = Hongaria G = Jerman
- Fi = Finlandia Fr = Perancis

Matriks jaraknya adalah sebagai berikut:

	<i>E</i>	<i>N</i>	<i>Da</i>	<i>Du</i>	<i>G</i>	<i>Fr</i>	<i>Sp</i>	<i>I</i>	<i>P</i>	<i>H</i>	<i>Fi</i>
<i>E</i>	0										
<i>N</i>	2	0									
<i>Da</i>	2	1	0								
<i>Du</i>	7	5	6	0							
<i>G</i>	6	4	5	5	0						
<i>Fr</i>	6	6	6	9	7	0					
<i>Sp</i>	6	6	5	9	7	2	0				
<i>I</i>	6	6	5	9	7	1	1	0			
<i>P</i>	7	7	6	10	8	5	3	4	0		
<i>H</i>	9	8	8	8	9	10	10	10	10	0	
<i>Fi</i>	9	9	9	9	9	9	9	9	9	8	0

Dari data, jarak minimum = 1 yaitu $d(DaN) = 1$, $d(Ifr) = 1$, $d(Isp) = 1$.

Karena $d(SpFr) = 2$; maka pada tingkat pertama yang dapat digabungkan adalah

kelompok I dan Fr atau kelompok I dan Sp . Dipilih kelompok I dan Fr bergabung membentuk kelompok (IFr) dengan jarak = 1.

Kemudian dihitung jarak antar obyek sebagai berikut:

$$d(IFr)E = \frac{d(IE) + d(FrE)}{N(IFr)NE} = \frac{6 + 6}{2.1} = 6$$

analog dengan jarak antara obyek (IFr) dengan obyek E , diperoleh:

$$\begin{aligned} d(IFr)N &= 6 & d(IFr)Da &= 11/2 & d(IFr)Du &= 9 & d(IFr)G &= 7 \\ d(IFr)Sp &= 3/2 & d(IFr)P &= 9/2 & d(IFr)H &= 10 & d(IFr)Fi &= 9 \end{aligned}$$

Data yang diperoleh, dimasukkan kedalam matriks jarak:

$$D = \{d_{ik}\} = \begin{matrix} & \begin{matrix} E & N & Da & Du & G & (IFr) & Sp & P & H & Fi \end{matrix} \\ \begin{matrix} E \\ N \\ Da \\ Du \\ G \\ (IFr) \\ Sp \\ P \\ H \\ Fi \end{matrix} & \begin{bmatrix} 0 & & & & & & & & & \\ 2 & 0 & & & & & & & & \\ 2 & 1 & 0 & & & & & & & \\ 7 & 5 & 6 & 0 & & & & & & \\ 6 & 4 & 5 & 5 & 0 & & & & & \\ 6 & 6 & 11/2 & 9 & 7 & 0 & & & & \\ 6 & 6 & 5 & 9 & 7 & 3/2 & 0 & & & \\ 7 & 7 & 6 & 10 & 8 & 9/2 & 3 & 0 & & \\ 9 & 8 & 8 & 8 & 9 & 10 & 10 & 10 & 0 & \\ 9 & 9 & 9 & 9 & 9 & 9 & 9 & 9 & 8 & 0 \end{bmatrix} \end{matrix}$$

Terbentuk kelompok (DaN) dengan jarak = 1. Perhitungan selanjutnya:

$$d(DaN)E = \frac{d(DaE) + d(NE)}{N(DaN)NE} = \frac{2 + 2}{2.1} = 2$$

Analog dengan jarak antara obyek (DaN) dengan obyek E , diperoleh:

$$\begin{aligned} d(DaN)Du &= 11/2 & d(DaN)G &= 9/2 & d(DaN)(IFr) &= 23/4 \\ d(DaN)Sp &= 11/2 & d(DaN)P &= 13/2 & d(DaN)H &= 8 & d(DaN)Fi &= 9 \end{aligned}$$

Hasil perhitungan dimasukkan kedalam matriks jarak:

$$D = \{d_{ik}\} = \begin{matrix} & E & (DaN) & Du & G & (IFr) & Sp & P & H & Fi \\ \begin{matrix} E \\ (DaN) \\ Du \\ G \\ (IFr) \\ Sp \\ P \\ H \\ Fi \end{matrix} & \begin{bmatrix} 0 \\ 2 & 0 \\ 7 & 11/2 & 0 \\ 6 & 9/2 & 5 & 0 \\ 6 & 23/4 & 9 & 7 & 0 \\ 6 & 11/2 & 9 & 7 & 3/2 & 0 \\ 7 & 13/2 & 10 & 8 & 9/2 & 3 & 0 \\ 9 & 8 & 8 & 9 & 10 & 10 & 10 & 0 \\ 9 & 9 & 9 & 9 & 9 & 9 & 9 & 8 & 0 \end{bmatrix} \end{matrix}$$

Terbentuk kelompok baru (*SpIFr*) dengan jarak = 3/2. Kemudian dihitung lagi,

$$d(SpIFr)E = \frac{d(SpE) + d(IE) + d(FrE)}{N(SpIFr)NE} = \frac{6 + 6 + 6}{3.1} = 6$$

Analog dengan jarak yang diperoleh antara obyek (*SpIFr*) dengan obyek *E* diperoleh,

$$\begin{aligned} d(SpIFr)(DaN) &= 34/6 & d(SpIFr)Du &= 9 & d(SpIFr)G &= 7 \\ d(SpIFr)P &= 4 & d(SpIFr)H &= 10 & d(SpIFr)Fi &= 9 \end{aligned}$$

Hasil perhitungan dimasukkan dalam matriks jarak :

$$D = \{d_{ik}\} = \begin{matrix} & E & (DaN) & Du & G & (SpIFr) & P & H & Fi \\ \begin{matrix} E \\ (DaN) \\ Du \\ G \\ (SpIFr) \\ P \\ H \\ Fi \end{matrix} & \begin{bmatrix} 0 \\ 2 & 0 \\ 7 & 11/2 & 0 \\ 6 & 9/2 & 5 & 0 \\ 6 & 34/6 & 9 & 7 & 0 \\ 7 & 13/2 & 10 & 8 & 4 & 0 \\ 9 & 8 & 8 & 9 & 10 & 10 & 0 \\ 9 & 9 & 9 & 9 & 9 & 9 & 8 & 0 \end{bmatrix} \end{matrix}$$

Terbentuk kelompok baru (*EdaN*) dengan jarak = 2. Kembali dihitung:

$$d(EDaN)Du = \frac{d(EDu) + d(DaDu) + d(NDu)}{N(EDaN)NDu} = \frac{7 + 6 + 5}{3.1} = 6$$

Analog dengan perhitungan jarak antara obyek (*EdaN*) dengan obyek *Du* diperoleh,

$$d(EDaN)G = 5 \qquad d(EDaN)(SpIFr) = 52/9 \qquad d(EDaN)P = 20/3$$

$$d(EDaN)H = 25/3 \qquad d(EDaN)Fi = 9$$

Hasil yang diperoleh dimasukkan kedalam matriks jarak:

$$D = \{d_{ik}\} = \begin{matrix} & \begin{matrix} (EDaN) & Du & G & (SpIFr) & P & H & Fi \end{matrix} \\ \begin{matrix} (EDaN) \\ Du \\ G \\ (SpIFr) \\ P \\ H \\ FI \end{matrix} & \begin{bmatrix} 0 & & & & & & \\ 6 & 0 & & & & & \\ 5 & 5 & 0 & & & & \\ 52/9 & 9 & 7 & 0 & & & \\ 20/3 & 10 & 8 & 4 & 0 & & \\ 25/3 & 8 & 9 & 10 & 10 & 0 & \\ 9 & 8 & 9 & 9 & 9 & 8 & 0 \end{bmatrix} \end{matrix}$$

Terbentuk kelompok baru (*PSpIFr*) dengan jarak = 4. Dilakukan perhitungan :

$$d(SpIFrP)(EDaN) = \frac{\begin{bmatrix} d(SpE) + d(SpDa) + d(SpN) + d(IE) + d(IDa) + d(IN) \\ + d(FrE) + d(FrDa) + d(FrN) + d(PE) + d(PDa) + d(PN) \end{bmatrix}}{N(SpIFrP)N(EDaN)}$$

$$= \frac{6 + 5 + 6 + 6 + 5 + 6 + 6 + 6 + 6 + 7 + 6 + 7}{4.3} = 6$$

Analog dengan jarak yang diperoleh antara obyek (*SpIFrP*) dengan obyek (*EdaN*) diperoleh,

$$d(SpIFrP)Du = 37/4 \qquad d(SpIFrP)G = 29/4$$

$$d(SpIFrP)H = 10 \qquad d(SpIFrP)Fi = 9$$

Hasil perhitungan dimasukkan kedalam matriks jarak :

$$D = \{d_{ik}\} = \begin{matrix} & \begin{matrix} (EDaN) & Du & (SpIFrP) & H & Fi \end{matrix} \\ \begin{matrix} (EDaN) \\ Du \\ G \\ (SpIFrP) \\ H \\ Fi \end{matrix} & \begin{bmatrix} 0 & & & & \\ 6 & 0 & & & \\ 5 & 5 & 0 & & \\ 6 & 37/4 & 29/4 & 0 & \\ 25/3 & 8 & 9 & 10 & 0 \\ 9 & 8 & 9 & 9 & 8 \end{bmatrix} \end{matrix}$$

Dipilih kelompok baru yang terbentuk (*EDaNG*) dengan jarak = 5. Kemudian dilakukan perhitungan :

$$d(EDaNG)Du = \frac{d(EDu) + d(DaDu) + d(NDu) + d(GDu)}{N(EDaNG)NDu} = \frac{7 + 6 + 5 + 5}{4.1} = \frac{23}{4}$$

analog dengan jarak yang diperoleh antara obyek (*EDaNG*) dengan obyek *Du* diperoleh,

$$d(EDaNG)(SpIFr) = 101/16 \qquad d(EDaNG)H = 34/4 \qquad d(EDaNG)Fi = 9$$

Hasil perhitungan dimasukkan kedalam matriks jarak:

$$D = \{d_{ik}\} = \begin{matrix} & \begin{matrix} (EDaNG) & Du & (SpIFrP) & H & Fi \end{matrix} \\ \begin{matrix} (EDaNG) \\ Du \\ (SpIFrP) \\ H \\ Fi \end{matrix} & \begin{bmatrix} 0 & & & & \\ 23/4 & 0 & & & \\ 101/16 & 37/4 & 0 & & \\ 34/4 & 8 & 10 & 0 & \\ 9 & 8 & 9 & 8 & 0 \end{bmatrix} \end{matrix}$$

Terbentuk kelompok baru (*EDaNGDu*) dengan jarak = 23/4. Perhitungan:

$$d(EDaNGDu)(SpIFrP) = \frac{\left[\begin{aligned} &d(ESp) + d(EI) + d(EFr) + d(EP) + d(DaSp) + \\ &d(DaI) + d(DaFr) + d(DaP) + d(NSp) + d(NI) + \\ &d(NFr) + d(NP) + d(GSp) + d(GI) + d(GFr) + d(GP) \\ &+ d(DuSp) + d(DuI) + d(DuFr) + d(DuP) \end{aligned} \right]}{N(EDaNGDu)N(SpIFrP)}$$

$$= \frac{6+6+6+7+5+5+6+6+6+6+6+7+7+7+7+8+9+9+9+10}{5.4} = \frac{138}{20}$$

Analog dengan jarak yang diperoleh antara obyek (*EDaNGDu*) dan obyek (*SpIFrP*) diperoleh,

$$d(EDaNGDu)H = 42/5 \quad d(EDaNGDu)Fi = 9$$

Hasil Perhitungan dimasukkan kedalam matriks jarak:

$$D = \{d_{ik}\} = \begin{matrix} & \begin{matrix} (EDaNGDu) & (SpIFrP) & H & Fi \end{matrix} \\ \begin{matrix} (EDaNGDu) \\ (SpIFrP) \\ H \\ Fi \end{matrix} & \begin{bmatrix} 0 & & & \\ 138 / 20 & 0 & & \\ 42 / 5 & 10 & 0 & \\ 9 & 9 & 8 & 0 \end{bmatrix} \end{matrix}$$

Jarak terpendek adalah 138/20 dan terbentuk kelompok baru (*EDaNGDuSpIFrP*) kemudian dihitung jaraknya dengan kelompok yang lain.

$$d(EDaNGDuSpIFrP)H = \frac{\begin{aligned} &d(EH) + d(DaH) + d(NH) + d(GH) + d(SpH) \\ &+ d(IH) + d(FrH) + d(PH) \end{aligned}}{N(EDaNGDuSpIFrP)NH}$$

$$= \frac{9+8+8+9+8+10+10+10+10}{9.1} = \frac{82}{9}$$

Analog dengan jarak yang diperoleh antara obyek (*EDaNGDuSpIFrP*) dan obyek *H* diperoleh $d(EDaNGDuSpIFrP)H = 9$

Hasil perhitungan dimasukkan kedalam matriks jarak:

$(EDaNGDuSpIFrP) \ H \ Fi$

$$D = \{d_{ik}\} = \begin{matrix} (EDaNGDuSpI \ FrP) \\ H \\ Fi \end{matrix} \begin{bmatrix} 0 & & \\ 82/9 & 0 & \\ 9 & 8 & 0 \end{bmatrix}$$

Terbentuk kelompok baru (Hfi) dengan jarak = 8. Perhitungan:

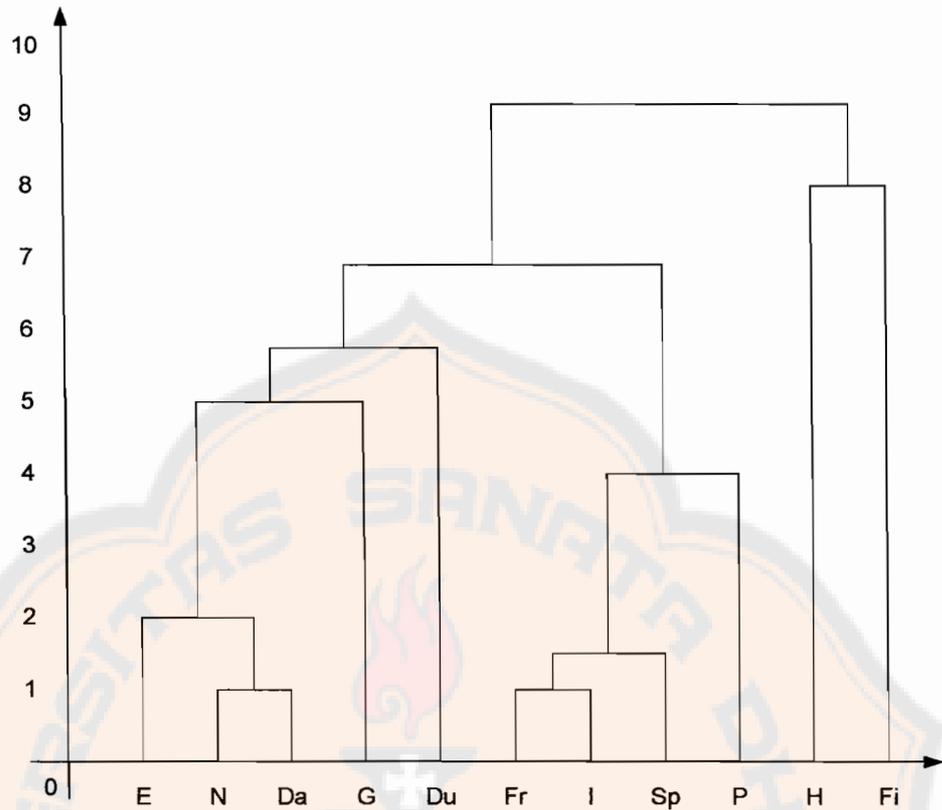
$$d(HFi)(EDaNGDuSpIFrP) = \frac{\begin{matrix} d(HE) + d(HDa) + d(HN) + d(HG) + d(HDu) + \\ d(HSp) + d(HI) + d(HFr) + d(HP) + d(FiE) + \\ d(FiDa) + d(FiN) + d(FiG) + d(FiDu) + d(FiSp) \\ + d(FiI) + d(FiFr) + d(FiP) \end{matrix}}{N(HFi)N(EDaNGDuSpIFrP)}$$

$$= \frac{9+8+8+9+8+10+10+10+10+9+9 + 9+9+9+9+9+9+9}{2.9} = \frac{163}{18}$$

Hasil ini dimasukkan kedalam matriks jarak:

$$D = \{d_{ik}\} = \begin{matrix} (EDaNGDuSpIFrP) \\ (HFi) \end{matrix} \begin{bmatrix} 0 & \\ 163/18 & 0 \end{bmatrix}$$

Untuk memperjelas proses pengelompokan dengan metode Average Linkage ini diberikan Dendogram berikut:



Dendrogram 3.4

3.2.3.1.4. Metode Centroid

Dalam metode centroid, jarak antara dua kelompok adalah jarak antara centroid kedua kelompok tersebut. Centroid kelompok adalah nilai tengah pada variabel observasi dalam kelompok variat. Dalam metode centroid, setiap waktu individu digabungkan, centroid baru dihitung.

Centroid kelompok R adalah $\bar{x}(R) = (\bar{x}_1(R), \bar{x}_2(R), \dots, \bar{x}_p(R))$ yang mempunyai

koordinat f sebagai berikut: $\bar{x}_f(R) = \frac{1}{|R|} \sum_{i \in R} x_{if} \dots \dots \dots (1)$

$|R|$ adalah banyaknya obyek dalam kelompok R . x_{if} adalah pengukuran ke f pada obyek ke i variabel x_i .

dimana $f = 1, \dots, p$ dan $i = 1, \dots, n$. Jarak Euclidean kuadrat antara dua obyek adalah juga bentuk kuadrat perbedaan kedua obyek tersebut.

$$d^2(u, v) = \|u - v\|^2 = (u - v) \cdot (u - v)$$

$$= u \cdot u - u \cdot v - u \cdot v + v \cdot v = \|u\|^2 + \|v\|^2 - 2u \cdot v \dots \dots \dots (2)$$

dimana \cdot adalah dot produk vektor

$$u \cdot v = (u_1, u_2, \dots, u_p) \cdot (v_1, v_2, \dots, v_p) = u_1v_1 + u_2v_2 + \dots + u_pv_p$$

Ketika menggabungkan kelompok A dan kelompok B untuk membentuk kelompok R , centroid R dapat ditulis sebagai fungsi dari A dan B :

$$\bar{x}(R) = \frac{|A|}{|R|} \bar{x}(A) + \frac{|B|}{|R|} \bar{x}(B) \dots \dots \dots (3)$$

dimana $|R| = |A| + |B|$. Identitas berikut akan digunakan lebih lanjut:

$$|A| \|\bar{x}(A) - \bar{x}(R)\|^2 + |B| \|\bar{x}(B) - \bar{x}(R)\|^2$$

$$= |A| \left\| \bar{x}(A) - \left(\frac{|A|}{|R|} \bar{x}(A) + \frac{|B|}{|R|} \bar{x}(B) \right) \right\|^2 + |B| \left\| \bar{x}(B) - \left(\frac{|A|}{|R|} \bar{x}(A) + \frac{|B|}{|R|} \bar{x}(B) \right) \right\|^2$$

$$= |A| \left\| \bar{x}(A) - \frac{|A|}{|R|} \bar{x}(A) - \frac{|B|}{|R|} \bar{x}(B) \right\|^2 + |B| \left\| \bar{x}(B) - \frac{|B|}{|R|} \bar{x}(B) - \frac{|A|}{|R|} \bar{x}(A) \right\|^2$$

$$\begin{aligned}
 &= |A| \left\| \left(\frac{|R|-|A|}{|R|} \right) \bar{x}(A) - \frac{|B|}{|R|} \bar{x}(B) \right\|^2 + |B| \left\| \left(\frac{|R|-|B|}{|R|} \right) \bar{x}(B) - \frac{|A|}{|R|} \bar{x}(A) \right\|^2 \\
 &= |A| \left\| \frac{|B|}{|R|} \bar{x}(A) - \frac{|B|}{|R|} \bar{x}(B) \right\|^2 + |B| \left\| \frac{|A|}{|R|} \bar{x}(B) - \frac{|A|}{|R|} \bar{x}(A) \right\|^2 \\
 &= \frac{|A||B|^2}{|R|^2} \|\bar{x}(A) - \bar{x}(B)\|^2 + \frac{|B||A|^2}{|R|^2} \|\bar{x}(A) - \bar{x}(B)\|^2 \\
 &= \frac{|A||B|^2 + |B||A|^2}{|R|^2} \|\bar{x}(A) - \bar{x}(B)\|^2 \\
 &= \frac{|A||B|(|B| + |A|)}{|R|^2} \|\bar{x}(A) - \bar{x}(B)\|^2 \\
 &= \frac{|A||B||R|}{|R|^2} \|\bar{x}(A) - \bar{x}(B)\|^2 \\
 &= \frac{|A||B|}{|R|} \|\bar{x}(A) - \bar{x}(B)\|^2 \dots\dots\dots(4)
 \end{aligned}$$

Dalam metode centroid (disebut juga *Centroid Linkage* atau *Centroid Sorting*), disimilaritas antara dua kelompok didefinisikan sebagai jarak Euclidean antara centroid kedua kelompok:

$$d(R,Q) = \|\bar{x}(R) - \bar{x}(Q)\| \dots\dots\dots(5)$$

persamaan (5) dapat dituliskan dalam bentuk berikut:

$$d^2(R,Q) = \frac{|A|}{|R|} d^2(A,Q) + \frac{|B|}{|R|} d^2(B,Q) - \frac{|A||B|}{|R|^2} d^2(A,B) \dots\dots\dots(6)$$

persamaan (6) mengandung bentuk kuadrat, sehingga nilai akhir $d(R,Q)$ adalah akar kuadrat sisi sebelah kanan persamaan (6).

Bukti:

$$\begin{aligned}
 & \frac{|A|}{|R|} d^2(A, Q) + \frac{|B|}{|R|} d^2(B, Q) - \frac{|A||B|}{|R|^2} d^2(A, B) \\
 &= \frac{|A|}{|R|} \|\bar{x}(A) - \bar{x}(Q)\|^2 + \frac{|B|}{|R|} \|\bar{x}(B) - \bar{x}(Q)\|^2 - \frac{|A||B|}{|R|^2} \|\bar{x}(B) - \bar{x}(Q)\|^2 \\
 &= \frac{|A|}{|R|} \left\| (\bar{x}(A) - \bar{x}(R)) + (\bar{x}(R) - \bar{x}(Q)) \right\|^2 + \frac{|B|}{|R|} \left\| (\bar{x}(B) - \bar{x}(R)) + (\bar{x}(R) - \bar{x}(Q)) \right\|^2 \\
 &\quad - \left(\frac{|A|}{|R|} \|\bar{x}(A) - \bar{x}(R)\|^2 + \frac{|B|}{|R|} \|\bar{x}(B) - \bar{x}(R)\|^2 \right) \\
 &= \frac{|A|}{|R|} \left\{ \|\bar{x}(A) - \bar{x}(R)\|^2 + \|\bar{x}(R) - \bar{x}(Q)\|^2 + 2(\bar{x}(A) - \bar{x}(R))(\bar{x}(R) - \bar{x}(Q)) \right\} \\
 &\quad + \frac{|B|}{|R|} \left\{ \|\bar{x}(B) - \bar{x}(R)\|^2 + \|\bar{x}(R) - \bar{x}(Q)\|^2 + 2(\bar{x}(B) - \bar{x}(R))(\bar{x}(R) - \bar{x}(Q)) \right\} \\
 &\quad - \frac{|A|}{|R|} \|\bar{x}(A) - \bar{x}(R)\|^2 - \frac{|B|}{|R|} \|\bar{x}(B) - \bar{x}(R)\|^2 \\
 &= \frac{|A|}{|R|} \left\{ \|\bar{x}(A) - \bar{x}(R)\|^2 + \|\bar{x}(R) - \bar{x}(Q)\|^2 + 2(\bar{x}(A) - \bar{x}(R))(\bar{x}(R) - \bar{x}(Q)) - \|\bar{x}(A) - \bar{x}(R)\|^2 \right\} \\
 &\quad + \frac{|B|}{|R|} \left\{ \|\bar{x}(B) - \bar{x}(R)\|^2 + \|\bar{x}(R) - \bar{x}(Q)\|^2 + 2(\bar{x}(B) - \bar{x}(R))(\bar{x}(R) - \bar{x}(Q)) - \|\bar{x}(B) - \bar{x}(R)\|^2 \right\} \\
 &= \frac{|A|}{|R|} \left\{ \|\bar{x}(R) - \bar{x}(Q)\|^2 + 2(\bar{x}(A) - \bar{x}(R))(\bar{x}(R) - \bar{x}(Q)) \right\} \\
 &\quad + \frac{|B|}{|R|} \left\{ \|\bar{x}(R) - \bar{x}(Q)\|^2 + 2(\bar{x}(B) - \bar{x}(R))(\bar{x}(R) - \bar{x}(Q)) \right\} \\
 &= \frac{|A|}{|R|} \|\bar{x}(R) - \bar{x}(Q)\|^2 + 2 \frac{|A|}{|R|} (\bar{x}(A) - \bar{x}(R))(\bar{x}(R) - \bar{x}(Q)) \\
 &\quad + \frac{|B|}{|R|} \|\bar{x}(R) - \bar{x}(Q)\|^2 + 2 \frac{|B|}{|R|} (\bar{x}(B) - \bar{x}(R))(\bar{x}(R) - \bar{x}(Q))
 \end{aligned}$$

$$\begin{aligned}
 &= \frac{|A|+|B|}{|R|} \|\bar{x}(R) - \bar{x}(Q)\|^2 + 2(\bar{x}(R) - \bar{x}(Q)) \left\{ \frac{|A|}{|R|} (\bar{x}(A) - \bar{x}(R)) + \frac{|B|}{|R|} (\bar{x}(B) - \bar{x}(R)) \right\} \\
 &= \|\bar{x}(R) - \bar{x}(Q)\|^2 + 2(\bar{x}(R) - \bar{x}(Q)) \left\{ \frac{|A|}{|R|} \bar{x}(A) + \frac{|B|}{|R|} \bar{x}(B) - \frac{|A|+|B|}{|R|} \bar{x}(R) \right\} \\
 &= \|\bar{x}(R) - \bar{x}(Q)\|^2 + 2(\bar{x}(R) - \bar{x}(Q)) (\bar{x}(R) - \bar{x}(R)) \\
 &= \|\bar{x}(R) - \bar{x}(Q)\|^2
 \end{aligned}$$

jadi, $d(R,Q) = \sqrt{\frac{|A|}{|R|} d^2(A,Q) + \frac{|B|}{|R|} d^2(B,Q) - \frac{|A||B|}{|R|^2} d^2(A,B)}$ □

Contoh 3.4

Diberikan data sebagai berikut:

obyek	Centroid		
	\bar{X}	\bar{Y}	\bar{Z}
A	4	8	2
B	7	3	5
C	3	2	9
D	4	6	7

Tabel 3.3

Dengan menggunakan metode centroid, dissimilaritas antara dua kelompok dihitung dengan menggunakan jarak Euclidean:

$$d(A,B) = \sqrt{(4-7)^2 + (8-3)^2 + (2-5)^2} = \sqrt{43} = 6,557$$

Analog dengan perhitungan dissimilaritas kelompok A dan B, diperoleh:

$$d(A,C) = 9,274 \quad d(A,D) = 5,385 \quad d(B,C) = 5,745$$

$$d(B,D) = 4,69 \quad d(C,D) = 4,582$$

Hasil yang diperoleh dimasukkan kedalam matriks jarak:

$$D = \{d_{ik}\} = \begin{matrix} & A & B & C & D \\ \begin{matrix} A \\ B \\ C \\ D \end{matrix} & \begin{bmatrix} 0 & & & \\ 6,557 & 0 & & \\ 9,274 & 5,745 & 0 & \\ 5,385 & 4,69 & 4,582 & 0 \end{bmatrix} \end{matrix}$$

Terbentuk kelompok (CD) dengan jarak = 4,582 kemudian dihitung kembali centroidnya.

Obyek	Centroid		
	\bar{X}	\bar{Y}	\bar{Z}
A	4	8	2
B	7	3	5
(CD)	$\frac{3+4}{2} = \frac{7}{2}$	$\frac{2+6}{2} = 4$	$\frac{9+7}{2} = 8$

Tabel 3.4

Dihitung kembali:

$$d(A, B) = \sqrt{(4-7)^2 + (8-3)^2 + (2-5)^2} = \sqrt{43} = 6,557$$

$$d(A, (CD)) = \sqrt{(4-3,5)^2 + (8-4)^2 + (2-8)^2} = \sqrt{52,25} = 7,228$$

$$d(B, (CD)) = \sqrt{(7-3,5)^2 + (3-4)^2 + (5-8)^2} = \sqrt{22,25} = 4,717$$

Hasil yang diperoleh dimasukkan kedalam matriks jarak:

$$D = \{d_{ik}\} = \begin{matrix} & A & B & (CD) \\ \begin{matrix} A \\ B \\ (CD) \end{matrix} & \begin{bmatrix} 0 & & \\ 6,557 & 0 & \\ 7,228 & 4,717 & 0 \end{bmatrix} \end{matrix}$$

Terbentuk kelompok baru (BCD) dengan jarak = 4,717 Untuk itu dilakukan perhitungan ulang pada centroid.

Obyek	Centroid		
	\bar{X}	\bar{Y}	\bar{Z}
A	4	8	2
(BCD)	$\frac{7+4+3}{3} = \frac{14}{3}$	$\frac{3+2+6}{3} = \frac{11}{3}$	$\frac{5+9+7}{3} = 7$

Tabel 3.5

Dihitung jaraknya:

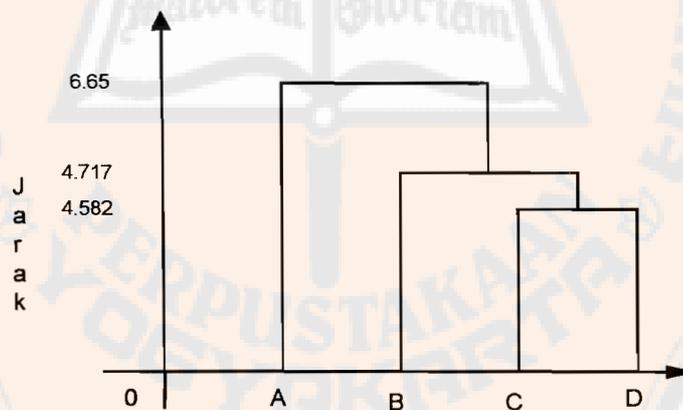
$$d(A, (BCD)) = \sqrt{\left(4 - \frac{14}{3}\right)^2 + \left(8 - \frac{11}{3}\right)^2 + (2 - 7)^2} = \sqrt{44,222} = 6,65$$

Dimasukkan ke dalam matriks jarak:

$$D = \{d_{ik}\} = \begin{matrix} & A & (BCD) \\ \begin{matrix} A \\ (BCD) \end{matrix} & \begin{bmatrix} 0 & \\ 6,65 & 0 \end{bmatrix} \end{matrix}$$

Terbentuk kelompok (ABCD) dengan jarak = 6,65

Untuk memperjelas proses pada metode Centroid diberikan dendrogram sebagai berikut:



Dendrogram 3.5

3.2.3.1.5. Metode Ward

Ward (1963) mengusulkan suatu metode yang membentuk kelompok berdasarkan pada hilangnya informasi hasil dari pengelompokan individu – individu kedalam kelompok yang diukur dengan total jumlah kuadrat galat (JKG) setiap observasi dari rata – rata kelompok yang ada.

Seperti metode centroid, metode Ward menggunakan jarak Euclidean antara centroid kedua kelompok untuk menghitung disimilaritas antara dua kelompok, tetapi jarak Euclidean tersebut dikalikan dengan suatu faktor pengali sebagai berikut:

$$d^2(R,Q) = \frac{2|R||Q|}{|R|+|Q|} \|\bar{x}(R) - \bar{x}(Q)\|^2 \dots\dots\dots(7)$$

Nilai $d(R,Q)$ adalah akar kuadrat persamaan (7).

Jumlah Kuadrat Galat (JKG)

Jumlah kuadrat galat (*JKG*) adalah istilah lain dari Error sum of squares (ESS). *JKG* digunakan sebagai ukuran kedekatan antar kelompok. *JKG* kelompok *C* didefinisikan sebagai jumlah jarak Euclidean kuadrat antara obyek – obyek dalam kelompok dan centroidnya:

$$JKG(C) = \sum_{i \in C} \|x_i - \bar{x}(C)\|^2 \dots\dots\dots(8)$$

Pada setiap tingkat, dipertimbangkan total *JKG* sebagai jumlah semua *JKG(C)* seluruh kelompok *C* dalam tingkat tersebut. Dengan menggabungkan kelompok *A* dan *B* untuk membentuk kelompok *R* pada tingkat selanjutnya, nilai *JKG* semakin besar. Besarnya perubahan (*increment*) dari *JKG* adalah:

$$\begin{aligned}
 \Delta JKG &= JKG(R) - JKG(A) - JKG(B) \\
 &= \sum_{i \in R} \|x_i - \bar{x}(R)\|^2 - \sum_{i \in A} \|x_i - \bar{x}(A)\|^2 - \sum_{i \in B} \|x_i - \bar{x}(B)\|^2 \\
 &= \sum_{i \in A} \|x_i - \bar{x}(R)\|^2 - \sum_{i \in B} \|x_i - \bar{x}(R)\|^2 - \sum_{i \in A} \|x_i - \bar{x}(A)\|^2 - \sum_{i \in B} \|x_i - \bar{x}(B)\|^2 \\
 &= \sum_{i \in A} \left(\|x_i - \bar{x}(R)\|^2 - \|x_i - \bar{x}(A)\|^2 \right) + \sum_{i \in B} \left(\|x_i - \bar{x}(R)\|^2 - \|x_i - \bar{x}(B)\|^2 \right)
 \end{aligned}$$

dengan menggunakan $\|u\|^2 - \|v\|^2 = u \cdot u - v \cdot v = u \cdot u + (u \cdot v - u \cdot v)$

$$\begin{aligned}
 &= u \cdot u + u \cdot v - u \cdot v - v \cdot v \\
 &= (u + v) \cdot (u - v)
 \end{aligned}$$

diperoleh,

$$\begin{aligned}
 &= \sum_{i \in A} \left[(x_i - \bar{x}(R) + x_i - \bar{x}(A)) \cdot (x_i - \bar{x}(R) - x_i + \bar{x}(A)) \right] + \\
 &\quad \sum_{i \in B} \left[(x_i - \bar{x}(R) + x_i - \bar{x}(B)) \cdot (x_i - \bar{x}(R) - x_i + \bar{x}(B)) \right] \\
 &= \left[\sum_{i \in A} (x_i - \bar{x}(R) + x_i - \bar{x}(A)) \right] \cdot (\bar{x}(A) - \bar{x}(R)) + \\
 &\quad \left[\sum_{i \in B} (x_i - \bar{x}(R) + x_i - \bar{x}(B)) \right] \cdot (\bar{x}(B) - \bar{x}(R))
 \end{aligned}$$

Diketahui bahwa $\sum_{i \in A} x_i = |A| \bar{x}(A)$ dan menggunakan persamaan (3), diperoleh:

$$\begin{aligned}
 &= |A| \left[(\bar{x}(A) - \bar{x}(R) + \bar{x}(A) - \bar{x}(A)) \cdot (\bar{x}(A) - \bar{x}(R)) \right] + \\
 &\quad |B| \left[(\bar{x}(B) - \bar{x}(R) + \bar{x}(B) - \bar{x}(B)) \cdot (\bar{x}(B) - \bar{x}(R)) \right] \\
 &= |A| (\bar{x}(A) - \bar{x}(R)) \cdot (\bar{x}(A) - \bar{x}(R)) + |B| (\bar{x}(B) - \bar{x}(R)) \cdot (\bar{x}(B) - \bar{x}(R)) \\
 &= |A| \|\bar{x}(A) - \bar{x}(R)\|^2 + |B| \|\bar{x}(B) - \bar{x}(R)\|^2
 \end{aligned}$$

$$= \frac{|A||B|}{|R|} \|\bar{x}(A) - \bar{x}(B)\|^2$$

Jadi $\Delta JKG = \frac{|A||B|}{|R|} \|\bar{x}(A) - \bar{x}(B)\|^2$

dengan menggunakan persamaan (7) dan mengganti kelompok R dan Q masing – masing dengan kelompok A dan B diperoleh:

$$\Delta JKG = \frac{1}{2} d^2(A, B) \text{ atau } d(A, B) = \sqrt{2\Delta JKG} \dots\dots\dots(9)$$

Algoritma yang dikembangkan oleh Ward menggunakan prosedur pengelompokan bertingkat. Proses pengelompokan dimulai dengan K kelompok dengan tiap kelompok terdiri dari satu obyek. Kelompok pertama dibentuk dengan memilih dua dari K kelompok tersebut, kemudian disatukan dan menghasilkan pasangan dengan nilai JKG yang paling kecil. K kelompok menjadi $K - 1$ kelompok, kemudian dibentuk kelompok kedua dengan memilih dua dari $K - 1$ kelompok tersebut yang akan disatukan dengan meminimalkan kenaikan nilai JKG . Dengan prosedur pengelompokan hierarki, K kelompok secara sistematis berkurang dari K kelompok menjadi $K - 1$ kelompok kemudian menjadi $K - 2$ kelompok sampai akhirnya menjadi 1 kelompok. Pada setiap tingkat (dari K sampai 1) nilai JKG diperoleh. Perubahan nilai JKG dari tingkat ke tingkat memberi petunjuk penting untuk menyatakan banyaknya kelompok ‘alami’ dari K kelompok. Metode Ward mungkin tidak menghasilkan nilai JKG yang optimal untuk banyaknya kelompok tertentu. Nilai JKG yang tidak optimal dapat terjadi dalam keadaan profil kelompok alami tidak jelas.

Berdasarkan hasil persamaan (9), metode Ward digunakan untuk menggabungkan kelompok A dan B dengan disimilaritas $d(A,B)$ minimal.

Persamaan (7) dapat dituliskan dalam bentuk berikut:

$$d^2(R,Q) = \frac{|A|+|Q|}{|R|+|Q|} d^2(A,Q) + \frac{|B|+|Q|}{|R|+|Q|} d^2(B,Q) - \frac{|Q|}{|R|+|Q|} d^2(A,B) \dots\dots\dots(10)$$

Untuk membuktikannya, digunakan persamaan (3).

Bukti:

$$\begin{aligned} d^2(R,Q) &= \frac{2|R||Q|}{|R|+|Q|} \|\bar{x}(R) - \bar{x}(Q)\|^2 \\ &= \frac{2|R||Q|}{|R|+|Q|} \left\| \left(\frac{|A|}{|R|} \bar{x}(A) + \frac{|B|}{|R|} \bar{x}(B) \right) - \bar{x}(Q) \right\|^2 \\ &= \frac{2|R||Q|}{|R|+|Q|} \left\| \left(\frac{|A|}{|R|} \bar{x}(A) - \bar{x}(Q) \right) + \left(\frac{|B|}{|R|} \bar{x}(B) - \bar{x}(Q) \right) \right\|^2 \\ &= \frac{2|R||Q|}{|R|+|Q|} \left\{ \frac{|A|^2}{|R|^2} \|\bar{x}(A) - \bar{x}(Q)\|^2 + \frac{|B|^2}{|R|^2} \|\bar{x}(B) - \bar{x}(Q)\|^2 \right. \\ &\quad \left. + 2 \frac{|A||B|}{|R||R|} (\bar{x}(A) - \bar{x}(Q)) \cdot (\bar{x}(B) - \bar{x}(Q)) \right\} \end{aligned}$$

dengan menggunakan persamaan $2u \cdot v = \|u\|^2 + \|v\|^2 - \|u - v\|^2$, diperoleh:

$$\begin{aligned} &= \frac{2|Q|}{(|R|+|Q|)|R|} \left\{ |A|^2 \|\bar{x}(A) - \bar{x}(Q)\|^2 + |B|^2 \|\bar{x}(B) - \bar{x}(Q)\|^2 \right. \\ &\quad \left. + |A||B| (\|\bar{x}(A) - \bar{x}(Q)\|^2 + \|\bar{x}(B) - \bar{x}(Q)\|^2 - \|\bar{x}(A) - \bar{x}(B)\|^2) \right\} \\ &= \frac{2|Q||A|}{|R|+|Q|} \frac{|A|+|B|}{|R|} \|\bar{x}(A) - \bar{x}(Q)\|^2 + \frac{2|Q||B|}{|R|+|Q|} \frac{|B|+|A|}{|R|} \|\bar{x}(B) - \bar{x}(Q)\|^2 \\ &\quad - \frac{|Q|}{|R|+|Q|} \frac{2|A||B|}{|R|} \|\bar{x}(A) - \bar{x}(B)\|^2 \end{aligned}$$

dengan memfaktorkan $\frac{|A|+|B|}{|R|}$, diperoleh:

$$\begin{aligned}
 &= \frac{2|A||Q|}{|A|+|Q|} \frac{|A|+|Q|}{|R|+|Q|} \|\bar{x}(A) - \bar{x}(Q)\|^2 + \frac{2|B||Q|}{|B|+|Q|} \frac{|B|+|Q|}{|R|+|Q|} \|\bar{x}(B) - \bar{x}(Q)\|^2 \\
 &\quad - \frac{2|A||B|}{|R|} \frac{|Q|}{|R|+|Q|} \|\bar{x}(A) - \bar{x}(B)\|^2 \\
 &= \frac{|A|+|Q|}{|R|+|Q|} d^2(A,Q) + \frac{|B|+|Q|}{|R|+|Q|} d^2(B,Q) - \frac{|Q|}{|R|+|Q|} d^2(A,B)
 \end{aligned}$$

Jadi terbukti bahwa

$$d^2(R,Q) = \frac{|A|+|Q|}{|R|+|Q|} d^2(A,Q) + \frac{|B|+|Q|}{|R|+|Q|} d^2(B,Q) - \frac{|Q|}{|R|+|Q|} d^2(A,B) \quad \square$$

Contoh 3.4

Diberikan data sifat yang dimiliki oleh 5 orang, sebagai berikut:

Orang	Nilai Sifat
A	2
B	5
C	9
D	10
E	15

Tabel 3.6

Dengan Metode Ward, akan dihitung ukuran kedekatan antar obyek sebagai berikut:

Perhitungan:

- Tingkat Pertama

Dicari centroid kelompok:

Karena setiap kelompok terdiri atas obyek tunggal, maka centroid tiap kelompok sama dengan nilai tiap kelompok tersebut. Nilai centroid dimasukkan kedalam matriks berikut:

Kelompok	Centroid
A	2
B	5
C	9
D	10
E	15

Tabel 3.7

Kemudian dihitung JKG masing – masing kelompok:

$$JKG(A) = \|2 - 2\|^2 = 0$$

Analog dengan $JKG(A)$, diperoleh:

$$JKG(B) = 0 \quad JKG(C) = 0 \quad JKG(D) = 0 \quad JKG(E) = 0.$$

Jadi pada tingkat pertama, JKG tiap kelompok = 0.

Kemudian dihitung jarak tiap kelompok, yaitu:

$$d(A) = \sqrt{2JKG(A)} = \sqrt{2 \times 0} = 0$$

Analog dengan $d(A)$, diperoleh:

$$d(B) = 0 \quad d(C) = 0 \quad d(D) = 0 \quad d(E) = 0$$

- Tingkat kedua

Centroid kelompok dihitung kembali:

Kelompok	Centroid	Kelompok	Centroid
AB	$(2 + 5) / 2 = 3,5$	BD	$(5 + 10) / 2 = 7,5$
AC	$(2 + 9) / 2 = 5,5$	BE	$(5 + 15) / 2 = 10$

AD	$(2 + 10) / 2 = 6$	CD	$(9 + 10) / 2 = 9,5$
AE	$(2 + 15) / 2 = 8,5$	CE	$(9 + 15) / 2 = 12$
BC	$(5 + 9) / 2 = 7$	DE	$(10 + 15) / 2 = 12,5$

Tabel 3.8

Kemudian dihitung *JKG* masing – masing kelompok, sebagai berikut:

$$JKG(AB) = \|2 - 3,5\|^2 + \|5 - 3,5\|^2 = 4,5$$

Analog dengan perhitungan *JKG* (*AB*), diperoleh *JKG* kelompok yang lain yaitu:

$$JKG(AC) = 24,5 \quad JKG(AD) = 32 \quad JKG(AE) = 84,5$$

$$JKG(BC) = 8 \quad JKG(BD) = 12,5 \quad JKG(BE) = 50$$

$$JKG(CD) = 0,5 \quad JKG(CE) = 18 \quad JKG(DE) = 12,5$$

Kemudian dihitung jarak antar kelompok:

$$d(A, B) = \sqrt{2JKG(AB)} = \sqrt{2 \times 4,5} = \sqrt{9} = 3$$

Analog dengan jarak kelompok *A* dan *B*, diperoleh:

$$d(A, C) = 7 \quad d(A, D) = 8 \quad d(A, E) = 13$$

$$d(B, C) = 4 \quad d(B, D) = 5 \quad d(B, E) = 10$$

$$d(C, D) = 1 \quad d(C, E) = 6 \quad d(D, E) = 5$$

Dengan metode Ward dipilih jarak antar kelompok dengan nilai paling kecil. Dari hasil perhitungan, terbentuk kelompok baru (*CD*) dengan jarak = 1.

- Tingkat ketiga

Centroid kelompok dihitung kembali:

Kelompok	Centroid
AB	$(2 + 5) / 2 = 3,5$
ACD	$(2 + 9 + 10) / 3 = 7$
AE	$(2 + 15) / 2 = 8,5$

BCD	$(5 + 9 + 10) / 3 = 8$
BE	$(5 + 15) / 2 = 10$
CDE	$(9 + 10 + 15) = 34 / 3$

Tabel 3.9

Kemudian dihitung *JKG* masing – masing kelompok sebagai berikut:

$$JKG(ACD) = \|2 - 7\|^2 + \|9 - 7\|^2 + \|10 - 7\|^2 = 38$$

Analog dengan perhitungan *JKG* (*ACD*), diperoleh *JKG* kelompok lain sebagai berikut:

$$JKG(AB) = 4,5 \qquad JKG(AE) = 84,5 \qquad JKG(BCD) = 14$$

$$JKG(BE) = 50 \qquad JKG(CDE) = 20,67$$

Kemudian dihitung jarak antar kelompok:

$$d(C, DE) = \sqrt{2JKG(CDE)} = \sqrt{2 \times 20,67} = \sqrt{41,34} = 6,43$$

Analog dengan perhitungan diatas, diperoleh:

$$d(A,B) = 3 \qquad d(A,CD) = 8,72 \qquad d(A,E) = 13$$

$$d(B,CD) = 5,29 \qquad d(B,E) = 10$$

Dengan metode Ward dipilih jarak antar kelompok dengan nilai paling kecil. Dari perhitungan diatas, terbentuk kelompok baru (*AB*) dengan jarak = 3.

- Tingkat empat

Dilakukan perhitungan ulang Centroid kelompok, yaitu:

Kelompok	Centroid
ABCD	$(2 + 5 + 9 + 10) / 4 = 6,5$
ABE	$(2 + 5 + 15) / 3 = 22 / 3$
CDE	$(9 + 10 + 15) / 3 = 34 / 3$

Tabel 3.10

Dihitung JKG kelompok:

$$JKG(ABCD) = \|2 - 6,5\|^2 + \|5 - 6,5\|^2 + \|9 - 6,5\|^2 + \|10 - 6,5\|^2 = 33$$

dan analog dengan $JKG(ABCD)$, diperoleh:

$$JKG(ABE) = 92,67 \quad JKG(CDE) = 20,67$$

Kemudian dihitung jarak antar dua kelompok, sebagai berikut:

$$d(AB,CD) = \sqrt{2JKG(ABCD)} = \sqrt{2 \times 33} = \sqrt{66} = 8,12$$

Analog dengan perhitungan diatas, diperoleh:

$$d(AB,E) = 13,61 \quad d(CD,E) = 6,43$$

Dengan metode Ward dipilih jarak antar kelompok dengan nilai paling kecil.

Sehingga terbentuk kelompok baru (CDE) dengan jarak = 6,43

- Tingkat Lima

Dihitung centroid kelompok sebagai berikut:

Kelompok	Centroid
ABCDE	$(2 + 5 + 9 + 10 + 15) / 5 = 8,2$

Tabel 3.11

Dihitung JKG kelompok, yaitu:

$$JKG(ABCDE) = \|2 - 8,2\|^2 + \|5 - 8,2\|^2 + \|9 - 8,2\|^2 + \|10 - 8,2\|^2 + \|15 - 8,2\|^2 = 95,56$$

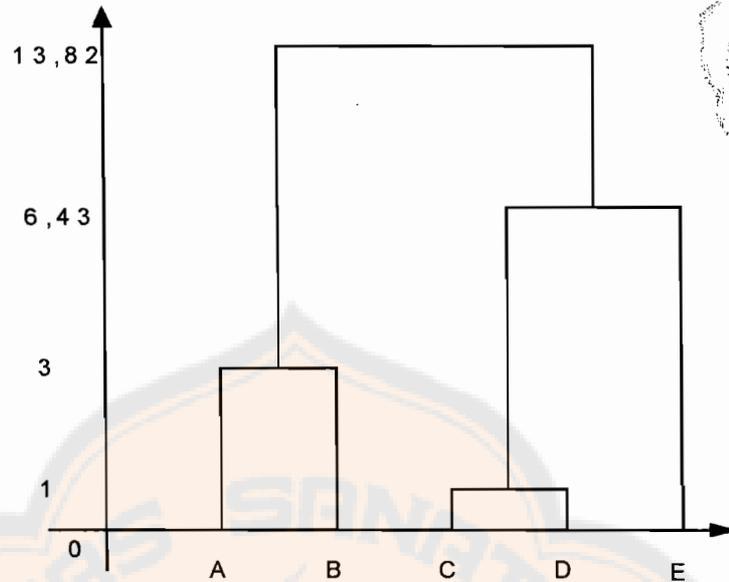
Kemudian dihitung jaraknya sebagaiberikut:

$$d(AB,CDE) = \sqrt{2JKG(ABCDE)} = \sqrt{2 \times 95,56} = \sqrt{191,12} = 13,82$$

Terbentuk kelompok $(ABCDE)$ dengan jarak = 13,82.

Proses pengelompokan tersebut, akan diperjelas dengan menggunakan dendogram

berikut:



Dendogram 3.6

3.2.3.2. Metode Non Hierarki

Berbeda dengan metode hierarki, metode pengelompokan nonhierarki tidak meliputi proses pengelompokan bertingkat seperti pohon. Metode nonhierarki menentukan obyek - obyek yang masuk kedalam kelompok - kelompok sedangkan banyaknya kelompok yang akan dibentuk ditentukan terlebih dahulu (dua kelompok, tiga kelompok atau yang lainnya). Banyaknya kelompok yang akan dibentuk tergantung pada tujuan penelitian serta subyektifitas peneliti. Setelah banyaknya kelompok ditentukan, proses pengelompokan dilakukan tanpa melalui proses hierarki.

Dalam prosedur pengelompokan nonhierarki, akan dibicarakan satu metode yang populer yaitu metode *K - mean*.

3.2.3.2.1. Metode *K - mean*

Mac Queen (1967) mengusulkan istilah *K- mean* untuk menggambarkan algoritma yang menandai setiap obyek pada kelompok yang mempunyai centroid terdekat.

Andaikan diketahui ada n obyek dan p pengukuran. $X(i, j)$ menyatakan nilai obyek i pada variabel j . Dimana $i = 1, 2, \dots, n$ dan $j = 1, 2, \dots, p$. Andaikan $P(n,K)$ merupakan partisi yang dihasilkan setiap kali n obyek dialokasikan ke salah satu kelompok $1, 2, \dots, K$. Mean variabel j dalam kelompok l dinyatakan dengan $\bar{X}(l, j)$ dan banyaknya obyek dalam kelompok l dinyatakan dengan $n(l)$. Jarak antara obyek i dan kelompok l dinyatakan sebagai berikut:

$$D(i, l) = \left(\sum_{j=1}^p [X(i, j) - \bar{X}(l, j)]^2 \right)^{\frac{1}{2}}$$

Komponen galat partisi didefinisikan sebagai berikut:

$$JKG[P(n, K)] = \sum_{i=1}^n D[i, l(i)]^2$$

Dimana $l(i)$ adalah kelompok yang mengandung obyek i dan $D(i, l(i))$ adalah jarak Euclidean kuadrat antara obyek i dan mean dari kelompok yang mengandung obyek i . Prosedur pengelompokan adalah mencari partisi dengan komponen galat (JKG) kecil dengan memindahkan obyek dari satu kelompok pada kelompok lain sampai tidak ada pemindahan obyek yang mengurangi nilai JKG .

Secara umum, proses pengelompokan dengan metode K - *mean* terdiri dari tiga langkah / tahap yaitu:

1. Membuat partisi obyek - obyek ke dalam K kelompok awal.
Sebetulnya, tidak ada ketentuan atau rumusan yang pasti mengenai banyaknya kelompok yang akan dibentuk. Hal tersebut tergantung pada tujuan penelitian serta subyektifitas dari peneliti sendiri.
2. Menandai dan mendaftarkan obyek - obyek yang mempunyai centroid terdekat. Jarak yang digunakan adalah jarak Euclidean tanpa melihat apakah obyek - obyek tersebut melalui proses standarisasi atau tidak. Menghitung ulang centroid untuk kelompok yang menerima obyek baru dan kelompok yang kehilangan obyek.
3. Ulangi langkah 2 sampai tidak ada obyek yang berpindah tempat.

Contoh 3.5

Diberikan observasi dua variabel X_1 dan X_2 untuk setiap obyek. Data diberikan dalam tabel berikut:

obyek	Observasi	
	X_1	X_2
A	5	3
B	-1	1
C	1	-2
D	-3	-2

Tabel 3.12

Berdasarkan observasi tersebut obyek – obyek dikelompokkan kedalam 2 kelompok ($K = 2$) sedemikian sehingga obyek - obyek dalam satu kelompok lebih dekat dari pada obyek - obyek dalam kelompok yang berbeda.

Langkah 1.

Dengan metode $K - mean$ dengan $K = 2$, obyek - obyek dibagi ke dalam dua kelompok awal. Misalnya (AB) dan (CD) . Kemudian dihitung koordinat centroid kelompok (\bar{X}_1, \bar{X}_2) . Hasil perhitungan dimasukkan kedalam matriks sebagai berikut:

kelompok	koordinat Centroid	
	\bar{X}_1	\bar{X}_2
(AB)	$\frac{5 + (-1)}{2} = 2$	$\frac{3 + 1}{2} = 2$
(CD)	$\frac{1 + (-3)}{2} = -1$	$\frac{(-2) + (-2)}{2} = -2$

Tabel 3.13

Langkah 2.

Menghitung jarak Euclidean untuk setiap obyek dari centroid kelompok dan menandai obyek dari kelompok dengan jarak terdekat. Jika sebuah obyek dipindah dari kelompok awal, maka centroid kelompok harus dihitung ulang sebelum proses dilanjutkan. Jarak dihitung dengan jarak Euclidean sebagai berikut:

$$d(A, (AB)) = \sqrt{(5 - 2)^2 + (3 - 2)^2} = \sqrt{10} = 3,16$$

Analog dengan $d(A, (AB))$, diperoleh:

$$d(A, (CD)) = 7,81$$

$$d(B, (AB)) = 3,16$$

$$d(B, (CD)) = 3$$

$$d(C, (AB)) = 4,12$$

$$d(C, (CD)) = 2$$

$$d(D, (AB)) = 6,4$$

$$d(D, (CD)) = 2$$

Hasil perhitungan dimasukkan kedalam matriks sebagai berikut:

	(AB)	(CD)
A	3,16	7,81
B	3,16	3
C	4,12	2
D	6,4	2

Tabel 3.14

Dari matriks jarak, dapat disimpulkan:

- ✓ A lebih dekat dengan kelompok (AB) daripada dengan kelompok (CD)
- ✓ B lebih dekat dengan kelompok (CD) daripada dengan kelompok (AB), sehingga terbentuk kelompok baru (BCD)
- ✓ C lebih dekat dengan kelompok (CD) daripada dengan kelompok (AB)
- ✓ D lebih dekat dengan kelompok (CD) daripada dengan kelompok (AB)

Akibatnya, kelompok B bergabung dengan kelompok (CD) membentuk kelompok baru (BCD). Maka koordinat centroid dihitung kembali.

Kelompok	Koordinat	Centroid
	\bar{X}_1	\bar{X}_2
A	5	3
(BCD)	$\frac{-1+1-3}{3} = -1$	$\frac{1-2-2}{3} = -1$

Tabel 3.15

Dihitung jarak setiapobyek dari centroid kelompok:

$$d(A, B) = \sqrt{(5+1)^2 + (3-1)^2} = \sqrt{40} = 6,32$$

Analog dengan $d(A, B)$, diperoleh:

$$d(A, C) = 6,4$$

$$d(A, D) = 9,43$$

$$d(A, (BCD)) = 7,21$$

$$d(B, (BCD)) = 2$$

$$d(C, (BCD)) = 2,24$$

$$d(D, (BCD)) = 2,24$$

Hasil perhitungan dimasukkan kedalam matriks sebagai berikut:

Kelompok	Obyek			
	A	B	C	D
A	0	6,32	6,4	9,43
(BCD)	7,21	2	2,24	2,24

Tabel 3.16

Dari perhitungan dapat disimpulkan, tidak terbentuk kelompok baru karena:

- ✓ *A* lebih dekat dengan kelompok *A* daripada dengan kelompok (*BCD*)
- ✓ *B* lebih dekat dengan kelompok (*BCD*) daripada dengan kelompok *A*
- ✓ *C* lebih dekat dengan kelompok (*BCD*) daripada dengan kelompok *A*
- ✓ *D* lebih dekat dengan kelompok (*BCD*) daripada dengan kelompok *A*

Oleh karena itu, tidak terbentuk kelompok baru sehingga proses berhenti.

Akhirnya, $K = 2$ kelompok tersebut adalah kelompok *A* dan kelompok (*BCD*).

Contoh 3.6

Algoritma *K - mean* digambarkan dengan tiga kandungan gizi yang ada dalam enam macam ikan. Data ditunjukkan dalam tabel berikut:

Macam ikan	Energi	Lemak	Kalsium
Ikan Air Tawar (IA)	5	9	20
Ikan Merah (IM)	6	11	2
Ikan Salem (SL)	4	5	20
Ikan Sarden (SD)	6	9	46
Ikan Tongkol (TN)	5	7	1
Udang (U)	3	1	12

Sumber: Hartigan (1975)

Tabel 3.17

Elemen dalam matriks dapat dinyatakan dengan $A(i, j)$ dengan $1 \leq i \leq 6$ dan $1 \leq j \leq 3$ masing – masing menyatakan banyaknya obyek dan banyaknya variabel.

Langkah membentuk kelompok berdasar algoritma $K - mean$ adalah sebagai berikut:

Langkah 1:

Jika jumlah elemen baris dinyatakan dengan $Sum(i)$, maka kelompok awal dapat dibentuk dengan mempertimbangkan obyek i sebagai bagian dari kelompok j , dan j adalah bagian dari $K[Sum(i) - Min] / (Max - Min) + 1$. Max dan Min masing – masing adalah nilai maksimum dan nilai minimum dari $Sum(i)$.

Nilai $Sum(i)$ ditunjukkan dalam tabel berikut:

Macam ikan	Energi	Lemak	Kalsium	Sum(i)
Ikan Air Tawar (IA)	5	9	20	34
Ikan Merah (IM)	6	11	2	19
Ikan Salem (SL)	4	5	20	29
Ikan Sarden (SD)	6	9	46	61
Ikan Tongkol (TN)	5	7	1	13
Udang (U)	3	1	12	16

Tabel 3.18

Andaikan K adalah banyaknya kelompok.

Dari tabel diatas, diperoleh: $Max = 61$ dan $Min = 13$. Dengan metode $K - mean$ dengan $K = 3$, obyek – obyek dibagi kedalam tiga kelompok awal misalnya:

Kelompok 1 (IM, TN, U)

Kelompok 2 (IA, SL)

Kelompok 3 (SD)

Langkah 2:

Menghitung $B(l,j)$ yaitu mean seluruh obyek variabel j dalam kelompok l .

Hasil perhitungan ditunjukkan dalam matriks berikut:

Kelompok	Energi	Lemak	Kalsium
(IM, TN, U)	14/3	19/3	5
(IA, SL)	9/2	7	20
(SD)	6	9	46

Tabel 3.19

Langkah 3:

Jarak antara obyek i dan kelompok l diberikan oleh:

$$D(i,l) = \left[\sum_{j=1}^3 \{A(i,j) - B(l,j)\}^2 \right]^{\frac{1}{2}} \text{ dan galat partisi didefinisikan sebagai berikut:}$$

$$JKG[P(n,K)] = \sum_{i=1}^n D(i,l(i))^2 \text{ dengan } l(i) \text{ adalah kelompok yang mengandung}$$

obyek i

Galat kelompok awal adalah sebagai berikut:

$$\begin{aligned} JKG(P(6,3)) &= D(1,2)^2 + D(2,1)^2 + D(3,2)^2 + D(4,3)^2 + D(5,1)^2 + D(6,1)^2 \\ &= \left(5 - \frac{9}{2}\right)^2 + (9 - 7)^2 + (20 - 20)^2 + \left(6 - \frac{14}{3}\right)^2 + \left(11 - \frac{19}{3}\right)^2 + (2 - 5)^2 \\ &\quad + \left(4 - \frac{9}{2}\right)^2 + (5 - 7)^2 + (20 - 20)^2 + (6 - 6)^2 + (9 - 9)^2 + (46 - 46)^2 \\ &\quad + \left(5 - \frac{14}{3}\right)^2 + \left(7 - \frac{19}{3}\right)^2 + (1 - 5)^2 + \left(3 - \frac{14}{3}\right)^2 + \left(1 - \frac{19}{3}\right)^2 + (12 - 5)^2 \\ &= 137,805 \end{aligned}$$

Langkah 4:

Memeriksa apakah ada perpindahan beberapa obyek dari satu kelompok ke kelompok lain yang mengurangi nilai JKG . Menghitung setiap obyek dengan rumus sebagai berikut:

$$R_{l(i),l} = \frac{n(l)D(i,l)^2}{n(l)+1} - \frac{n(l(i))D(i,l(i))^2}{n(l(i))-1}$$

dengan $n(l)$ = banyaknya obyek dalam kelompok l

$l(i)$ = kelompok yang mengandung obyek i

Perhitungan pada obyek pertama adalah sebagai berikut:

$$D(1,1)^2 = \left(5 - \frac{14}{3}\right)^2 + \left(9 - \frac{19}{3}\right)^2 + (20 - 5)^2 = 232,22$$

$$D(1,2)^2 = \left(5 - \frac{9}{2}\right)^2 + (9 - 7)^2 + (20 - 20)^2 = 4,25$$

$$D(1,3)^2 = (5 - 6)^2 + (9 - 9)^2 + (20 - 46)^2 = 677$$

$$R_{2(1),1} = \frac{3}{4}(232,22) - 2(4,25) > 0$$

$$R_{2(1),3} = \frac{1}{2}(677) - 2(4,25) > 0$$

Terlihat bahwa perpindahan obyek 1 dari kelompok 2 ke kelompok 1 atau kelompok 3 mengakibatkan nilai JKG bertambah sehingga obyek 1 tetap dalam kelompok 2. Analog dengan perhitungan pada obyek 1, diperoleh obyek 2, 3, 4, dan 5 tetap dalam kelompok awal. Perhitungan untuk obyek 6 ditunjukkan sebagai berikut:

$$D(6,1)^2 = \left(3 - \frac{14}{3}\right)^2 + \left(1 - \frac{19}{3}\right)^2 + (12 - 5)^2 = 80,21$$

$$D(6,2)^2 = \left(3 - \frac{9}{2}\right)^2 + (1-7)^2 + (12-20)^2 = 102,25$$

$$D(6,3)^2 = (3-6)^2 + (1-9)^2 + (12-46)^2 = 1229$$

$$R_{1(6),2} = \frac{2}{3}(102,25) - \frac{3}{2}(80,21) = -52,15$$

$$R_{1(6),3} = \frac{1}{2}(1229) - \frac{3}{2}(80,21) > 0$$

Jika obyek 6 berpindah dari kelompok 1 ke kelompok 2 maka terjadi pengurangan nilai *JKG* sebesar 52,15 sehingga diperoleh:

$$JKG[P'(n, K)] = 137,805 - 52,15 = 85,655$$

dimana $P'(n, K)$ adalah partisi (IM, TN) (IA, SL, U) (SD)

Langkah 5:

Matriks mean $B(l, j)$ yang baru adalah sebagai berikut:

Kelompok	Energi	Lemak	Kalsium
(IM, TN)	11/2	9	3/2
(IA, SL, U)	4	5	52/3
(SD)	6	9	46

Tabel 3.20

Perhitungan seperti langkah 4, diperoleh $R_{1(i),l}$ positif dan tidak ada perpindahan obyek dari satu kelompok ke kelompok lain yang mengurangi nilai *JKG*.

Penyelesaian akhir adalah sebagai berikut:

Kelompok 1 (IM, TN)

Kelompok 2 (IA, SL, U)

Kelompok 3 (SD)

Disimpulkan bahwa dari ketiga variabel (energi, lemak, kalsium), variabel kalsium mempunyai perbedaan yang besar diantara ketiga kelompok. Oleh karena itu, kalsium merupakan faktor yang mendominasi perbedaan antar kelompok bukan energi dan lemak. Kelompok 3 mempunyai kalsium paling tinggi yang besarnya = 46 dan kelompok 1 mempunyai kalsium paling rendah yang besarnya = 3/2 dan besarnya kalsium pada kelompok 2 = 52/3.

Setelah terbentuk kelompok baru dengan menggunakan salah satu metode, baik dengan metode hierarki ataupun metode nonhierarki, maka langkah selanjutnya adalah melakukan interpretasi terhadap kelompok yang telah terbentuk.

3.2.4. Interpretasi Kelompok

Interpretasi kelompok meliputi pemeriksaan setiap kelompok dalam kelompok variat dengan memberi nama atau label yang menggambarkan kelompok alami secara tepat.

3.2.5. Membuat Profil Kelompok

Membuat profil meliputi penggambaran karakteristik setiap kelompok yang digunakan untuk membedakan karakteristik tersebut pada dimensi yang relevan. Prosedur dimulai setelah kelompok – kelompok diidentifikasi atau di beri label. Peneliti menggunakan data dari variabel – variabel yang tidak dilibatkan dalam pengelompokan awal.

Membuat profil dan interpretasi kelompok lebih dari sekedar menggambarkan kelompok tersebut. Pertama, memberi arti pada taksiran hubungan antar kelompok dengan menggunakan teori sebelumnya atau pengalaman praktis. Kedua, profil kelompok memberikan jalan untuk membuat taksiran signifikan yang praktis. Untuk menjelaskan proses ini, diberikan contoh kasus penelitian segmentasi pasar produk shampoo Clear (Pratiwi dan Dharmmesta, 2000).

Contoh 3.7

Dilakukan penelitian tentang segmentasi pasar produk shampoo Clear.

Data yang diperlukan meliputi:

1. Karakteristik responden.

Data karakteristik konsumen digunakan untuk mengetahui profil konsumen shampoo Clear yang terdiri dari:

- Variabel demografis: jenis kelamin, usia, pendidikan dan pengeluaran dalam sebulan.
- Variabel geografis: tempat tinggal responden.
- Variabel perilaku: produk shampoo yang pernah digunakan, status penggunaan shampoo Clear, dan intensitas pemakaian shampoo dalam satu minggu.
- Variabel psikografis: penggunaan jasa salon dalam merawat rambut.

2. Atribut produk digunakan oleh konsumen sebagai pertimbangan dalam memilih shampoo Clear. Atribut produk adalah hal – hal /sifat – sifat yang melekat pada produk yaitu: Manfaat produk efektif bagi rambut dan kulit kepala, formula produk efektif dan berteknologi tinggi (Zinc System), harga produk yang terjangkau, kemudahan mendapatkan produk di toko ketika diperlukan, citra produk yang baik, desain kemasan produk menarik, iklan produk yang menarik.

3. Data – data dasar segmentasi pasar.

Data – data dasar segmentasi pasar digunakan untuk mendeskripsikan karakteristik konsumen secara lebih spesifik. Dasar segmentasi pasar yang digunakan adalah dasar segmentasi psikografis dan perilaku, khususnya menekankan pada variabel gaya hidup dan manfaat yang dicari.

Ruang lingkup penelitian adalah sebagai berikut:

1. Obyek penelitian adalah produk shampoo Clear.
2. Subyek penelitian adalah konsumen yang sedang menggunakan dan yang pernah menggunakan produk shampoo Clear.
3. Wilayah penelitian dilakukan di Daerah Istimewa Yogyakarta.
4. Dasar segmentasi pasar yang digunakan adalah psikografis dan perilaku, dengan penekanan pada variabel gaya hidup dan variabel manfaat yang dicari.

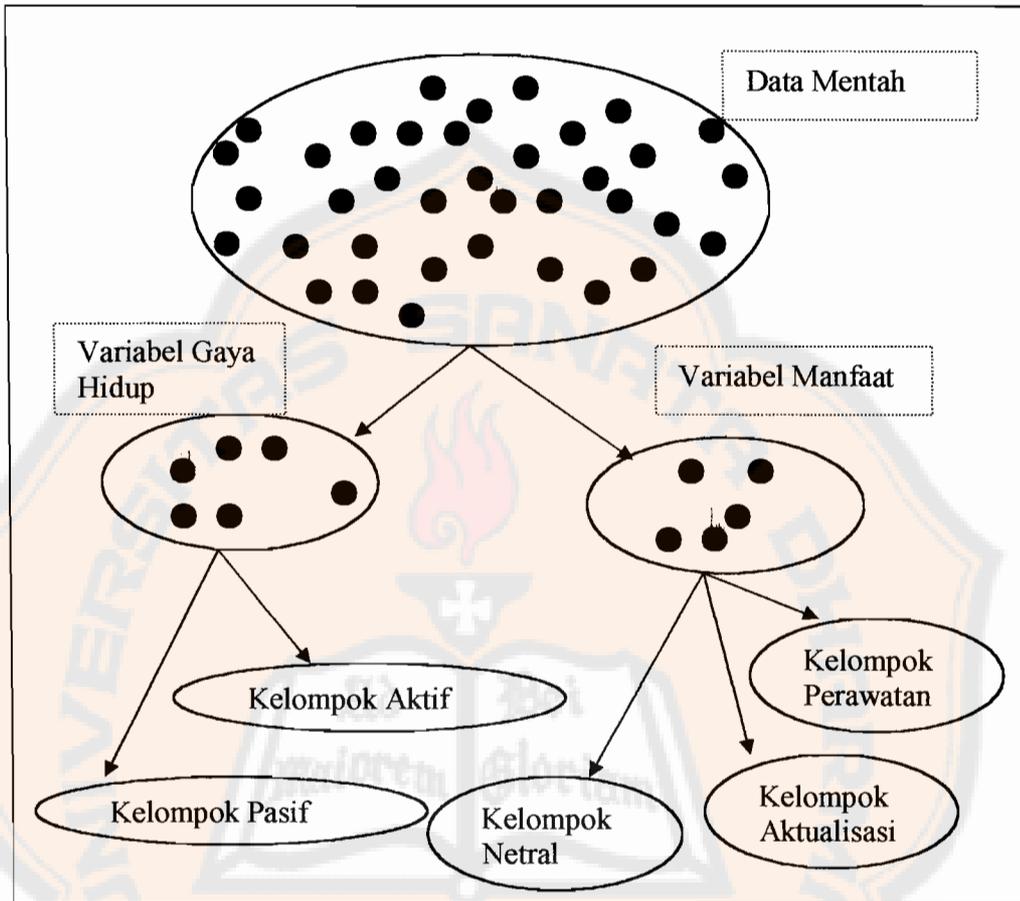
Langkah – langkah analisis data :

1. Melakukan analisis faktor untuk mereduksi variabel yang berkorelasi tinggi. Analisis faktor terhadap variabel gaya hidup menghasilkan 6 variabel sedangkan pada variabel manfaat menghasilkan 5 variabel. Analisis faktor tidak dibahas dalam skripsi ini.
2. Melakukan analisis kelompok untuk mengelompokkan konsumen shampoo Clear menjadi beberapa kelompok.
3. Interpretasi setiap kelompok.
4. Menganalisis profil kelompok
 - Untuk menentukan profil kelompok diperlukan atribut produk dan karakteristik konsumen.

Data diperoleh secara langsung dari sumber dengan menyebarkan kuesioner kepada responden. Sampel yang diambil dalam penelitian ini adalah 100 orang responden pengguna produk shampoo Clear. Variabel yang dianalisis dalam variabel gaya hidup sebanyak 18 (aktif organisasi, pejabat penting, kegiatan sosial, pekerja keras, kemampuan diri, tugas tepat waktu, menjadi pemimpin, penampilan eksklusif, penampilan menarik, mengikuti tren, obsesi jabatan tinggi, dunia bisnis menarik, pendidikan lebih tinggi, berinteraksi dengan orang, rumah rapi, memperhatikan kesehatan, mengikuti isu politik dan prestasi baik). Variabel yang dianalisis dalam variabel manfaat sebanyak 19 (citra diri, rambut subur, gengsi naik, akar rambut, rambut hitam, semangat kerja, pergaulan luas, percaya diri, rambut sehat, nilai ekonomis, hilang penat,

kualitas produk, segar kulit kepala, penampilan baik, sejuk kulit kepala, rambut harum, halus dan rapi, hilang ketombe, cegah ketombe).

Bagan Langkah – langkah Analisis Data:



Gambar 3.3

Interpretasi Kelompok:

Berdasarkan variabel gaya hidup, terbentuk dua kelompok yaitu kelompok 1 yang memfokuskan pada aktif berorganisasi, percaya diri, mengejar karier dan berprestasi. Oleh karena itu, kelompok 1 disebut kelompok aktif. Kelompok 2 tidak memiliki kecenderungan seperti pada kelompok 1. Oleh karena itu, kelompok 2 dapat disebut kelompok pasif.

Berdasarkan variabel manfaat, terbentuk tiga kelompok yaitu: kelompok 1 memiliki kecenderungan rendah terhadap semua variabel. Kelompok 2 fokus pada meningkatkan aktualisasi, menjaga penampilan dan mengatasi ketombe. Kelompok 3 fokus pada merawat rambut dan merawat kulit kepala. Berdasarkan ciri yang dimiliki, kelompok 1 disebut kelompok netral, kelompok 2 disebut kelompok aktualisasi dan kelompok 3 disebut kelompok perawatan.

Profil Kelompok

Karakteristik konsumen shampoo Clear adalah sebagai berikut:

- Berdasarkan variabel demografis: Konsumen shampoo Clear sebagian besar pria; berusia relatif muda (21-25 tahun); memiliki latar belakang pendidikan sarjana/ sederajat; taraf ekonominya menengah kebawah (pengeluaran dalam sebulan di bawah 250.000).
- Berdasarkan variabel geografis: Tempat tinggal konsumen saat ini di asrama/indekos.
- Berdasarkan variabel perilaku: Shampoo yang digunakan konsumen sebanyak 1-3 shampoo; sebagian besar konsumen sedang menggunakan shampoo Clear; konsumen memakai shampoo 3 kali/ minggu.
- Berdasarkan variabel gaya hidup: sebagian besar konsumen tidak pernah ke salon.

Atribut Produk Shampoo Clear:

Dari berbagai atribut produk shampoo Clear yang melekat, diperoleh 7 atribut yang penting dan mendapatkan respon dari konsumen. Atribut produk digunakan konsumen sebagai pertimbangan dalam memilih shampoo Clear. Atribut produk berdasarkan tingkat kepentingannya berdasarkan skor rata – rata konsumen adalah sebagai berikut:

1. Manfaat produk efektif bagi rambut dan kulit kepala.
2. Formula produk efektif dan berteknologi tinggi (Zinc System)
3. Harga produk yang terjangkau.
4. Kemudahan mendapatkan produk di toko ketika diperlukan.
5. Citra produk yang baik.
6. Desain kemasan produk menarik.
7. Iklan produk yang menarik.

Dengan mengetahui atribut produk menurut tingkat kepentingannya, perusahaan dapat melakukan inovasi produk dengan menonjolkan atribut – atribut utama yang dikehendaki oleh konsumen. Sehingga konsumen akan loyal terhadap produk tersebut karena karakteristik produk telah dikenal dengan baik dan mampu memenuhi kebutuhan serta keinginan konsumen.

Dengan mengetahui karakteristik konsumen, perusahaan dapat melakukan pemasaran didaerah yang kebanyakan asrama / indekos atau daerah sekitar perguruan tinggi.

PLAGIAT MERUPAKAN TINDAKAN TIDAK TERPUJI

BAB IV

APLIKASI ANALISIS KELOMPOK

Dalam bab ini, analisis kelompok diterapkan dalam bidang pendidikan dan bidang kesehatan. Pengolahan data dilakukan dengan program SPSS. Penggunaan program SPSS ini bertujuan untuk memudahkan dan mempercepat proses analisis.

4.1. Bidang Pendidikan

Data mengenai jumlah perguruan tinggi, jumlah mahasiswa dan jumlah tenaga edukatif negeri dan swasta dibawah lingkungan Departemen Pendidikan Nasional, menurut Propinsi 2000 / 2001 bersumber dari BPS (2001). Data diambil dari 26 propinsi yang ada di Indonesia.

Propinsi – propinsi tersebut akan dikelompokkan kedalam tiga kelompok berdasarkan variabel jumlah perguruan tinggi negeri (jptn), jumlah perguruan tinggi swasta (jpts), jumlah mahasiswa negeri (jmhsn), jumlah mahasiswa swasta (jmhss), jumlah tenaga edukatif negeri (jten), dan jumlah tenaga edukatif swasta (jtes). Data diberikan pada Lampiran 1 yang berisi tentang jumlah Perguruan Tinggi, Mahasiswa dan Tenaga Edukatif Negeri dan Swasta di bawah Lingkungan Departemen Pendidikan Nasional, menurut Provinsi 2000/2001 Langkah – langkah analisis data dengan SPSS dapat dilihat pada lampiran 2

Output:

Analisis Kelompok dengan Metode Hierarki

Case Processing Summary^a

Cases					
Valid		Missing		Total	
N	Percent	N	Percent	N	Percent
26	100.0%	0	.0%	26	100.0%

a. Squared Euclidean Distance used

Tabel 4.1

Output Proximity Matrix dapat dilihat pada Lampiran 3

Agglomeration Schedule

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	5	7	9.803E-03	0	0	4
2	18	26	1.452E-02	0	0	5
3	17	19	1.736E-02	0	0	7
4	5	24	2.639E-02	1	0	5
5	5	18	3.564E-02	4	2	8
6	15	22	3.571E-02	0	0	8
7	17	25	5.078E-02	3	0	12
8	5	15	6.980E-02	5	6	10
9	1	8	9.767E-02	0	0	13
10	4	5	.104	0	8	15
11	16	20	.172	0	0	12
12	16	17	.266	11	7	15
13	1	6	.268	9	0	18
14	14	21	.292	0	0	16
15	4	16	.530	10	12	18
16	3	14	.639	0	14	20
17	2	23	.720	0	0	19
18	1	4	.736	13	15	20
19	2	12	1.091	17	0	22
20	1	3	3.057	18	16	22
21	10	11	3.374	0	0	23
22	1	2	5.690	20	19	25
23	10	13	6.790	21	0	24
24	9	10	18.352	0	23	25
25	1	9	37.481	22	24	0

Tabel 4.2

Cluster Membership

Case	3 Clusters
1:Dista Aceh	1
2:Sumatera Utara	1
3:Sumatera Barat	1
4:Riau	1
5:Jambi	1
6:Sumatera Selatan	1
7:Bengkulu	1
8:Lampung	1
9:DKI Jakarta	2
10:Jawa Barat	3
11:Jawa Tengah	3
12:DI Yogyakarta	1
13:Jawa Timur	3
14:Bali	1
15:Nusa Tenggara Barat	1
16:Nusa Tenggara Timur	1
17:Kalimantan Barat	1
18:Kalimantan Tengah	1
19:Kalimantan Selatan	1
20:Kalimantan Timur	1
21:Sulawesi Utara	1
22:Sulawesi Tengah	1
23:Sulawesi Selatan	1
24:Sulawesi Tenggara	1
25:Maluku	1
26:Irian Jaya	1

Tabel 4.3

Dendogram dengan metode Average Linkage dapat dilihat pada Lampiran 4

Interpretasi:

Tabel 4.1 menunjukkan bahwa semua data telah diproses tanpa ada data yang hilang (missing).

Tabel 4.2 menunjukkan hasil proses pengelompokan dengan metode Average Linkage. Pada tahap (stage) 1 terbentuk kelompok dengan anggota

obyek 5 (propinsi Jambi) dan obyek 7 (propinsi Bengkulu). Pada kolom koefisien, terdapat nilai $9,803E-3$ atau $0,009803$ yang menyatakan jarak antara propinsi Jambi dan propinsi Bengkulu. Pada Next Stage terdapat nilai 2, berarti langkah pengelompokan selanjutnya adalah stage 2 atau baris 2 dengan penjelasan analog pada baris 1.

Tabel 4.3 adalah daftar propinsi anggota dalam kelompok yang dibentuk. Ditentukan 3 kelompok maka dengan melihat kolom 3 Clusters:

- Anggota kelompok 1 adalah propinsi dengan tanda 1 sebanyak 22 propinsi, yaitu: Dista Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, D.I. Yogyakarta, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Maluku dan Irian Jaya.
- Anggota kelompok 2 adalah propinsi dengan tanda 2 yaitu propinsi DKI Jakarta.
- Anggota kelompok 3 adalah propinsi dengan tanda 3 yaitu propinsi Jawa Barat, Jawa Tengah dan Jawa Timur.

Analisis Kelompok dengan metode Nonhierarki

Initial Cluster Centers

	Cluster		
	1	2	3
Zscore(JPTN)	-.87832	.49186	1.40532
Zscore(JPTS)	-.75521	2.64028	1.16519
Zscore(JMHSN)	-.65748	3.77393	1.29363
Zscore(JMHSS)	-.60304	3.18390	.83669
Zscore(JTEN)	-.78514	1.19720	1.76585
Zscore(JTES)	-.60876	3.41313	.58418

Tabel 4.4

Final Cluster Centers

	Cluster		
	1	2	3
Zscore(JPTN)	-.39756	.49186	1.17695
Zscore(JPTS)	-.51204	2.64028	1.18143
Zscore(JMHSN)	-.45867	3.77393	.82346
Zscore(JMHSS)	-.49505	3.18390	1.03701
Zscore(JTEN)	-.49627	1.19720	1.37199
Zscore(JTES)	-.48391	3.41313	.96353

Tabel 4.5

ANOVA

	Cluster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
Zscore(JPTN)	5.778	2	.585	23	9.885	.001
Zscore(JPTS)	10.164	2	.203	23	50.027	.000
Zscore(JMHSN)	11.154	2	.117	23	95.308	.000
Zscore(JMHSS)	10.623	2	.163	23	65.086	.000
Zscore(JTEN)	8.703	2	.330	23	26.363	.000
Zscore(JTES)	10.835	2	.145	23	74.811	.000

The F tests should be used only for descriptive purposes because the clusters have been chosen to maximize the differences among cases in different clusters. The observed significance levels are not corrected for this and thus cannot be interpreted as tests of the hypothesis that the cluster means are equal.

Tabel 4.6

Number of Cases in each Cluster

Cluster	1	19.000
	2	1.000
	3	6.000
Valid		26.000
Missing		.000

Tabel 4.7

Interpretasi:

Dari tabel 4.6 semakin besar nilai F dan nilai signifikannya di bawah 0,05 maka semakin besar pula perbedaan variabel tersebut pada tiga kelompok. Nilai F terbesar adalah 95,308 ada pada nilai variabel ZJMHSN dengan nilai Sig = 0,000 berarti variabel jumlah mahasiswa negeri sangat membedakan karakteristik ketiga kelompok. Variabel jumlah perguruan tinggi swasta, jumlah mahasiswa swasta, jumlah tenaga edukatif negeri dan jumlah tenaga edukatif swasta mempunyai nilai F yang besar dan nilai Sig = 0,000 sehingga variabel - variabel tersebut juga membedakan karakteristik ketiga kelompok.

Tabel 4.5 menunjukkan bahwa dari proses pengelompokkan terbentuk 3 kelompok dengan ciri yang berbeda satu sama lain. Sebelum menafsir isi setiap kelompok, terlebih dahulu diperlihatkan contoh penafsiran sebuah variabel, misalnya variabel ZJPTN.

Untuk menghitung rata – rata jumlah perguruan tinggi setiap kelompok adalah sebagai berikut: $\bar{X} = \mu + Z\sigma$

dimana \bar{X} : rata – rata jumlah perguruan tinggi dalam satu kelompok.

μ : rata – rata jumlah perguruan tinggi seluruhnya.

σ : Standar Deviasi

Z : nilai standarisasi yang di dapat pada SPSS

Dari data, diperoleh: $\mu = 2,92$ dan $\sigma = 2,19$ dengan demikian

- Rata – rata jumlah perguruan tinggi negeri pada kelompok 1 = $2,92 + (-0,39756 \times 2,19) = 2,05 < 2,92$ berarti di bawah rata – rata.
- Rata – rata jumlah perguruan tinggi negeri pada kelompok 2 = $2,92 + (0,49186 \times 2,19) = 3,99 > 2,92$ berarti di atas rata – rata.
- Rata – rata jumlah perguruan tinggi negeri pada kelompok 3 = $2,92 + (1,17695 \times 2,19) = 5,53 > 2,92$ berarti di atas rata – rata.

Analog dengan perhitungan rata – rata variabel jumlah perguruan tinggi negeri, diperoleh rata – rata variabel yang lain.

Ketiga kelompok yang terbenntuk adalah sebagai berikut:

- Dari tabel 4.6 diketahui kelompok 1 terdiri dari 19 propinsi, yaitu: Propinsi Dista Aceh, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Sulawesi Utara, Sulawesi Tengah, Sulawesi Tenggara, Maluku dan Irian Jaya.

Kelompok 1 merupakan propinsi dengan jumlah perguruan tinggi negeri rata – rata = 2. Jumlah perguruan tinggi swasta, jumlah mahasiswa negeri, jumlah mahasiswa swasta, jumlah tenaga edukatif negeri dan jumlah tenaga edukatif swasta di bawah rata – rata.

Dari ciri – ciri diatas, dapat disimpulkan bahwa kelompok 1 adalah kelompok propinsi yang jumlah perguruan tinggi, jumlah mahasiswa, dan

jumlah tenaga edukatif dibawah rata – rata, meliputi propinsi – propinsi di luar Pulau Jawa

- Dari tabel 4.6 diketahui bahwa kelompok 2 merupakan propinsi dengan jumlah perguruan tinggi negeri rata – rata 3 sampai 4, dan jumlah perguruan tinggi swasta, jumlah mahasiswa negeri, jumlah perguruan tinggi swasta, jumlah tenaga edukatif negeri dan jumlah tenaga edukatif swasta di atas rata – rata. Kelompok 2 adalah propinsi yang jumlah perguruan tinggi, jumlah mahasiswa dan jumlah tenaga edukatif di atas rata – rata dan relatif tinggi. Dapat dikatakan bahwa D.K.I Jakarta merupakan propinsi dengan fasilitas pendidikan yang baik.

- Dari tabel 4.6 diketahui kelompok 3 terdiri dari 5 propinsi, yaitu propinsi Sumatera Utara, Jawa Barat, Jawa Tengah, D.I Yogyakarta, Jawa Timur dan Sulawesi Selatan.

Kelompok 3 merupakan propinsi dengan jumlah perguruan tinggi negeri, jumlah perguruan tinggi swasta, jumlah mahasiswa negeri, jumlah mahasiswa swasta, jumlah tenaga edukatif negeri dan jumlah tenaga edukatif swasta di atas rata – rata.

Dari ciri – ciri diatas, dapat disimpulkan bahwa kelompok 3 adalah kelompok propinsi dengan jumlah perguruan tinggi, jumlah mahasiswa, dan jumlah tenaga edukatif di atas rata – rata, meliputi propinsi – propinsi di Pulau Jawa.



4.2. Bidang Kesehatan

Data mengenai jumlah tenaga kesehatan daerah menurut Propinsi 2000 / 2001 diambil dari 26 propinsi yang ada di Indonesia.

Propinsi - propinsi tersebut akan dikelompokkan kedalam tiga kelompok berdasarkan variabel Dokter, SKM, APT DIII, Farmasi, Akper, Epedemiologi, Analisis Lab, Tenaga Elektro dan Teknik, SPK, Asisten APT, dan Sanitarian. Data yang diperoleh diberikan pada Lampiran 5.

Data: Bidang_Kesehatan

Buka file **Bidang_Kesehatan**

Langkah – langkah Analisis Kelompok dengan SPSSv10

Langkah – langkah analisis pada Bidang_Kesehatan analog dengan langkah – langkah analisis pada Bidang_Pendidikan dengan mengubah variabel – variabel jumlah perguruan tinggi negeri, jumlah perguruan tinggi swasta, jumlah mahasiswa negeri, jumlah mahasiswa swasta, jumlah tenaga edukatif negeri dan jumlah tenaga edukatif swasta dengan variabel – variabel dokter, skm, aptdiii, farmasi, akper, epedemil, anallabs, telektek, bidan, spk, assapt dan sanitari.

Output:

Analisis Kelompok dengan Metode Hierarki

Case Processing Summary^a

Cases					
Valid		Missing		Total	
N	Percent	N	Percent	N	Percent
26	100.0%	0	.0%	26	100.0%

a. Squared Euclidean Distance used

Tabel 4.8

Output Proximity Matrix dapat dilihat pada Lampiran 6

Agglomeration Schedule

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	7	21	.372	0	0	5
2	18	24	.717	0	0	4
3	14	16	.820	0	0	9
4	17	18	1.187	0	2	9
5	7	15	1.225	1	0	11
6	4	8	1.286	0	0	7
7	4	23	1.559	6	0	15
8	19	25	1.708	0	0	11
9	14	17	1.939	3	4	14
10	1	5	2.311	0	0	12
11	7	19	2.411	5	8	14
12	1	20	2.543	10	0	13
13	1	6	4.002	12	0	16
14	7	14	4.263	11	9	15
15	4	7	4.781	7	14	16
16	1	4	6.461	13	15	17
17	1	3	7.344	16	0	18
18	1	26	8.134	17	0	19
19	1	22	9.246	18	0	21
20	10	11	15.450	0	0	22
21	1	12	15.715	19	0	25
22	2	10	19.948	0	20	23
23	2	9	29.946	22	0	24
24	2	13	37.308	23	0	25
25	1	2	58.095	21	24	0

Tabel 4.9

Cluster Membership

Case	3 Clusters
1:DI Aceh	1
2:Sumatera Utara	2
3:Sumatera Barat	1
4:Riau	1
5:Jambi	1
6:Sumatera Selatan	1
7:Bengkulu	1
8:Lampung	1
9:DKI Jakarta	2
10:Jawa Barat	2
11:Jawa Tengah	2
12:DI Yogyakarta	1
13:Jawa Timur	3
14:Kalimantan Barat	1
15:Kalimantan Tengah	1
16:Kalimantan Selatan	1
17:Kalimantan Timur	1
18:Sulawesi Utara	1
19:Sulawesi Tengah	1
20:Sulawesi Selatan	1
21:Sulawesi Tenggara	1
22:Bali	1
23:Nusa Tenggara Barat	1
24:Nusa Tenggara Timur	1
25:maluku	1
26:Irian Jaya	1

Tabel 4.10

Dendogram dengan metode Average Linkage dapat dilihat pada Lampiran 7

Interpretasi:

Tabel 4.8 menunjukkan bahwa semua data telah diproses tanpa ada data yang hilang (missing).

Tabel 4.9 menunjukkan hasil proses pengelompokan dengan metode Average Linkage. Pada tahap (stage) 1 terbentuk kelompok dengan anggota

obyek 7 (propinsi Bengkulu) dan obyek 21 (propinsi Sulawesi Tenggara). Pada kolom koefisien, terdapat nilai 0,372 yang menyatakan jarak antara propinsi Bengkulu dan propinsi Sulawesi Tenggara. Pada Next Stage terdapat nilai 2, berarti langkah pengelompokan selanjutnya adalah stage 2 atau baris 2 dengan penjelasan analog pada baris 1.

Tabel 4.10 adalah daftar propinsi anggota dalam kelompok yang dibentuk. Ditentukan 3 kelompok maka dengan melihat kolom 3 Clusters:

- Anggota kelompok 1 adalah propinsi dengan tanda 1 sebanyak 21 propinsi, yaitu: Dista Aceh, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, D.I. Yogyakarta, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Maluku dan Irian Jaya.
- Anggota kelompok 2 adalah propinsi dengan tanda 2 yaitu propinsi DKI Jakarta.
- Anggota kelompok 3 adalah propinsi dengan tanda 3 sebanyak 4 propinsi, yaitu: Sumatera Utara, D.K.I Jakarta, Jawa Barat dan Jawa Tengah.

Analisis dengan Metode Nonhierarki

Initial Cluster Centers

	Cluster		
	1	2	3
Zscore(DOKTER)	-.61825	1.55141	2.77951
Zscore(SKM)	-.65084	3.81322	.33540
Zscore(APTIII)	-.63029	2.34689	2.78853
Zscore(FARMASI)	-.21301	.43690	2.30185
Zscore(AKPER)	-.58807	1.87442	.85318
Zscore(EPEDEMIL)	-.91702	.42245	3.63719
Zscore(ANALLABS)	-1.28599	1.56181	.55874
Zscore(TELEKTEK)	-.51866	2.59685	3.25931
Zscore(BIDAN)	-.63119	3.55500	1.67557
Zscore(SPK)	-.43654	1.50800	1.58365
Zscore(ASSAPT)	-.49156	2.35359	2.29654
Zscore(SANITARI)	.57009	-1.88377	.26439

Tabel 4.11

Final Cluster Centers

	Cluster		
	1	2	3
Zscore(DOKTER)	-.36811	1.55141	2.18236
Zscore(SKM)	-.28985	3.81322	.85448
Zscore(APTIII)	-.37947	2.34689	2.00051
Zscore(FARMASI)	-.32347	.43690	2.22650
Zscore(AKPER)	-.35728	1.87442	1.99524
Zscore(EPEDEMIL)	-.25946	.42245	1.76192
Zscore(ANALLABS)	-.25200	1.56181	1.32738
Zscore(TELEKTEK)	-.37317	2.59685	1.87099
Zscore(BIDAN)	-.36589	3.55500	1.49816
Zscore(SPK)	-.29489	1.50800	1.65983
Zscore(ASSAPT)	-.36785	2.35359	1.91307
Zscore(SANITARI)	.00300	-1.88377	.60589

Tabel 4.12

ANOVA

	Cluster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
Zscore(DOKTER)	9.838	2	.231	23	42.502	.000
Zscore(SKM)	9.290	2	.279	23	33.277	.000
Zscore(APTIII)	10.341	2	.188	23	55.081	.000
Zscore(FARMASI)	8.682	2	.332	23	26.154	.000
Zscore(AKPER)	9.132	2	.293	23	31.185	.000
Zscore(EPEDEMIL)	5.486	2	.610	23	8.996	.001
Zscore(ANALLABS)	4.561	2	.690	23	6.607	.005
Zscore(TELEKTEK)	10.155	2	.204	23	49.790	.000
Zscore(BIDAN)	11.158	2	.117	23	95.643	.000
Zscore(SPK)	6.226	2	.546	23	11.412	.000
Zscore(ASSAPT)	9.748	2	.239	23	40.734	.000
Zscore(SANITARI)	2.325	2	.885	23	2.628	.094

The F tests should be used only for descriptive purposes because the clusters have been chosen to maximize the differences among cases in different clusters. The observed significance levels are not corrected for this and thus cannot be interpreted as tests of the hypothesis that the cluster means are equal.

Tabel 4.13

Number of Cases in each Cluster

Cluster	1	22.000
	2	1.000
	3	3.000
Valid		26.000
Missing		.000

Tabel 4.14

Interpretasi:

Dari tabel 4.13 semakin besar nilai F dan nilai signifikannya dibawah 0,05 maka semakin besar pula perbedaan variabel tersebut pada tiga kelompok. Nilai F terbesar adalah 95,643 ada pada variabel ZBIDAN dengan nilai Sig = 0,000 berarti variabel jumlah bidan sangat membedakan karakteristik ketiga kelompok. Variabel jumlah sanitari mempunyai nilai F yang kecil dan nilai Sig = 0,094 sehingga variabel sanitari tidak membedakan karakteristik ketiga kelompok.

Tabel 4.12 menunjukkan bahwa dari proses pengelompokan terbentuk 3 kelompok dengan ciri yang berbeda satu sama lain. Sebelum menafsir isi setiap kelompok, terlebih dahulu diperlihatkan contoh penafsiran sebuah variabel, misalnya variabel ZDOKTER.

Untuk menghitung rata – rata jumlah dokter pada setiap kelompok adalah sebagai berikut: $\bar{X} = \mu + Z\sigma$

dimana \bar{X} : rata – rata jumlah dokter dalam setiap kelompok

μ : rata – rata jumlah dokter seluruhnya

σ : Standar Deviasi

Z : nilai standarisasi yang didapat pada SPSS

Dari data Bidang_Kesehatan, diperoleh: $\mu = 1035,27$ dan $\sigma = 1134,27$ dengan demikian,

- Rata – rata jumlah perguruan tinggi negeri pada kelompok 1 = $1035,27 + (-0,36811 \times 1134,27) = 617,73 < 1035,27$ berarti di bawah rata – rata.
- Rata – rata jumlah perguruan tinggi negeri pada kelompok 2 = $1035,27 + (1,55141 \times 1134,27) = 2794,98 > 1035,27$ berarti di atas rata – rata.
- Rata – rata jumlah perguruan tinggi negeri pada kelompok 3 = $1035,27 + (2,18236 \times 1134,27) = 3510,65 > 1035,27$ berarti di atas rata – rata.

Analog dengan perhitungan rata – rata variabel jumlah dokter, diperoleh rata – rata variabel yang lain.

Ketiga kelompok yang terbentuk adalah sebagai berikut:

- Dari tabel 4.12 diketahui kelompok 1 terdiri dari 22 propinsi, yaitu: Propinsi Dista Aceh, Sumatera Barat, Sumatera Utara, D.I Yogyakarta,

Sulawesi Selatan, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Sulawesi Utara, Sulawesi Tengah, Sulawesi Tenggara, Maluku dan Irian Jaya.

Kelompok 1 merupakan propinsi dengan jumlah dokter, jumlah SKM, jumlah APT DIII, jumlah farmasi, jumlah akper, jumlah epedemiologi, jumlah analis lab, jumlah tenaga elektro dan teknik, jumlah bidan, jumlah SPK dan jumlah asisten APT di bawah rata – rata.

Dari ciri – ciri di atas, dapat disimpulkan bahwa kelompok 1 adalah kelompok propinsi – propinsi yang jumlah tenaga kesehatan daerah sebagian besar di bawah rata - rata. Propinsi pada kelompok 1 sebagian besar terletak di luar Pulau Jawa.

- Dari tabel 4.12 diketahui kelompok 2 merupakan propinsi dengan jumlah dokter, jumlah SKM, jumlah APT DIII, jumlah farmasi, jumlah akper, jumlah epedemiologi, jumlah analis lab, jumlah tenaga elektro dan teknik, jumlah bidan, jumlah SPK dan jumlah asisten APT diatas rata – rata. Jadi, propinsi Jawa Timur merupakan propinsi dengan jumlah tenaga kesehatan daerah di atas rata – rata.
- Dari tabel 4.12 diketahui kelompok 3 terdiri dari 3 propinsi, yaitu propinsi D.K.I Jakarta, Jawa Barat dan Jawa Tengah.

Kelompok 3 merupakan propinsi dengan jumlah dokter, jumlah SKM, jumlah APT DIII, jumlah farmasi, jumlah akper, jumlah epedemiologi,

jumlah analis lab, jumlah tenaga elektro dan teknik, jumlah bidan, jumlah SPK, jumlah asisten APT di atas rata – rata.

Dari ciri – ciri di atas, dapat disimpulkan bahwa kelompok 3 terdiri dari propinsi – propinsi dengan jumlah tenaga kesehatan daerah yang relatif tinggi, meliputi propinsi – propinsi yang terletak di Pulau Jawa.



PLAGIAT MERUPAKAN TINDAKAN TIDAK TERPUJI

BAB V

KESIMPULAN

Analisis kelompok adalah teknik pengelompokan multivariat yang maksud utamanya adalah mengelompokkan obyek berdasarkan karakteristik yang dimiliki. Hasil pengelompokan obyek menunjukkan homogenitas (similaritas) yang tinggi antar anggota dalam satu kelompok dan heterogenitas (perbedaan) yang tinggi antar kelompok yang satu dengan kelompok yang lain.

Dalam proses pengelompokan dengan analisis kelompok, diasumsikan tidak ada data yang hilang (missing). Proses dasar analisis kelompok adalah sebagai berikut:

1. Mengukur similaritas antar kelompok
2. Proses standarisasi
3. Memilih algoritma pengelompokan
4. Interpretasi kelompok
5. Membuat profil kelompok

Ada dua metode yang digunakan dalam analisis kelompok, yaitu metode hierarki dan metode nonhierarki. Dalam metode hierarki, terdapat lima metode populer yang digunakan yaitu:

1. Metode Single Linkage
2. Metode Complete Linkage
3. Metode Average Linkage
4. Metode Centroid
5. Metode Ward's

Dalam metode nonhierarki, yaitu: metode *K – mean*.

Analisis kelompok dapat ditindak lanjuti dengan analisis diskriminan yang tidak dibahas dalam skripsi ini. Oleh karena itu, perlu adanya pembahasan analisis diskriminan sebagai lanjutan analisis kelompok.



PLAGIAT MERUPAKAN TINDAKAN TIDAK TERPUJI

DAFTAR PUSTAKA

- Anderson, Black, Hair, Jr., and Tatham. (1998). *Multivariate Data Analysis*, 5th edition. Upper Saddle River, NJ: PrenticeHall, inc.
- Flury, B. (1998). *A first Course in Multivariate Statistics*, 5th edition. New York: John Willey and Sons.
- Hand, D.J. (1981). *Discrimination and Classification*. Chichester : John Willey and Sons.
- Kaufman, L., Rousseeuw, P. (1990). *Finding Group in Data*. New York: John Willey and Sons.
- Kerlinger, F. (1964). *Foundation of Behavioral Research Educational and Psychological Enquiry*. England: Holt, Rinehart and Winston, Inc.
- Mendenhall, W., Scheaffer, R. L., Wackerly, D.D. (1986). *Mathematical Statistics With Applications*. Boston: Duxbury Press.
- Pratiwi, P., Dharmmesta, B.S. *Segmentasi Pasar Produk Shampoo Clear*. Kajian Bisnis. Mei – Agustus 2000. No 20.
- Richard, A.J., Dean, W.W. (1992). *Applied Multivariate Statistical Analysis*, 3rd edition. USA: PrenticeHall, inc.
- Santosa, S. (2002). *Buku Latihan SPSS Statistik Multivariat*. Jakarta: PT. Elex Media Komputindo.
- Walpole, R. (1990). *Pengantar Statistik*, Edisi ke-3. Jakarta: PT Gramedia.
- <http://www.statsoftinc.com/textbook/esc.html>
- <http://www.lss.ucc.ac.uk/postGrad/quant/measurement.htm>
- <http://web.uccs.edu/thecker/SPSS/scalemens.htm>

PLAGIAT MERUPAKAN TINDAKAN TIDAK TERPUJI

LAMPIRAN



Lampiran 1: Data Bidang Pendidikan

Propinsi	Jumlah Perguruan Tinggi		Jumlah Mahasiswa		JmlTenaga Edukatif	
	Negeri	Swasta	Negeri	Swasta	Negeri	Swasta
Dista Aceh Sumatera Utara	2	42	20540	30796	1469	2593
Sumatera Barat	3	114	35554	128418	2543	10748
Riau	5	53	26016	49877	2512	5114
Jambi	1	31	13179	19276	910	1247
Sumatera Selatan	1	10	6540	6414	625	753
Bengkulu	2	73	18759	44230	1361	4170
Lampung	1	13	9616	6790	577	550
D.K.I Jakarta	2	49	18099	23892	1006	1798
Jawa Barat	4	254	174136	401408	4118	31824
Jawa Tengah	7	226	86368	204698	6391	18534
D.I Yogyakarta	6	148	80331	156584	5120	9969
Jawa Timur	3	80	47343	166555	3481	9312
Bali	10	221	88310	332648	6251	21769
NusaTenggara Barat	4	24	18959	17671	2209	2562
NusaTenggara Timur	1	18	9429	14572	771	2471
Kalimantan Barat	3	8	8466	14673	859	383
Kalimantan Tengah	2	18	13277	15630	883	1042
Kalimantan Selatan	1	19	8344	7251	606	932
Kalimantan Timur	2	17	9463	20604	919	1549
Sulawesi Utara	3	33	13976	18378	938	1642
Sulawesi Tengah	4	34	28923	11569	2991	2183
Sulawesi Selatan	1	14	11682	7428	922	1463
Sulawesi Tenggara	4	106	37390	75969	2770	7066
Maluku	1	12	13879	4939	589	590
Irian Jaya	2	29	8530	7882	880	727
Indonesia	1	25	9443	14000	519	864
	76	1671	816552	1802152	52220	141855

Sumber: Departemen Pendidikan Nasional

Lampiran 2:

Langkah – langkah Analisis Kelompok dengan SPSSv10

1. Buka file **Bidang _ Pendidikan**

Standarisasi data dengan Z - Score

- Dari menu **Analyze**, pilih **Descriptives Statistics**, lalu pilih **Descriptives...**
- Masukkan kotak **VARIABLE(S)**, variabel **jptn, jpns, jmhsn, jmhss, jten dan jtes**.
- Kemudian aktifkan pilihan **Save Standardized values as variables**.
- Tekan **OK** untuk proses.

Analisis Kelompok dengan Metode Hierarki:

- Dari menu **Analyze**, pilih submenu **Classify**, lalu pilih **Hierarchical Cluster...**
Pengisian:
 - **VARIABLES**. Sesuai kasus, masukkan semua variabel yakni **jptn, jpns, jmhsn, jmhss, jten dan jtes**.
 - **LABEL CASES BY**, masukkan variabel **Propinsi**.
 - **CLUSTER**, akan dilakukan pengelompokan berdasarkan propinsi, biarkan pilihan pada **cases**.
 - **DISPLAY** atau tampilan yang akan ditayangkan pada output, yang meliputi pilihan Statistik yang relevan dan Plot (grafik). Aktifkan **Statistics dan Plots**.

Lampiran 2 (lanjutan):

- Kemudian klik kotak Statistics...
 - Aktifkan kotak **Proximity matrix** untuk menampilkan jarak antar variabel.
 - Pada bagian CLUSTER MEMBERSHIP, klik mouse pada pilihan SINGLE SOLUTIONS ketik 3

Kemudian tekan CONTINUE untuk kembali ke menu utama.

- Kemudian klik kotak **Plots...**

Pada kotak dialog PLOTS berisi pilihan grafik yang akan ditampilkan, aktifkan pilihan **Dendogram**.

Pada kotak dialog ICICLE merupakan bagian yang tidak akan ditampilkan, pilih **None**.

- Kemudian klik kotak Method...

Pengisian :

- CLUSTER METHOD pilih **Between group linkage**.
- MEASURE buka kotak combo dan pilih **Squared Euclidean distance**.
- TRANSFORM VALUE buka kotak combo dan pilih **Z scores**.
- Kemudian tekan CONTINUE untuk kembali ke menu utama
- Tekan **OK** untuk proses data.

Analisis Kelompok dengan Metode K - means

- Dari menu **analyze**, pilih submenu **Classify**, lalu pilih **K - means Cluster...**

Lampiran 2 (Lanjutan):

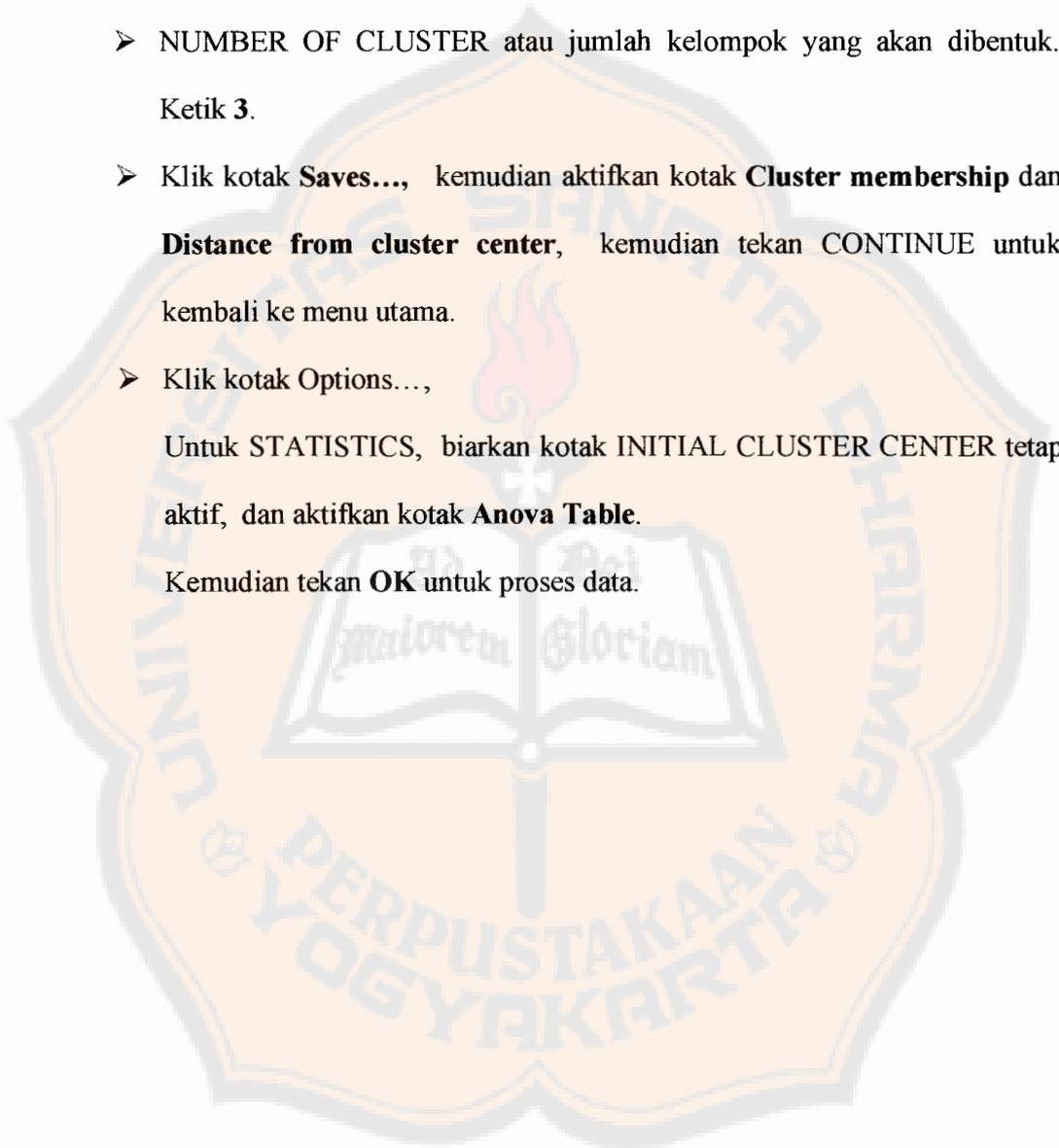
Pengisian:

- VARIABLES. Masukkan semua variabel dengan awalan z, yakni **zjptn, zjpts, zjmhsn, zjmhss, zjten dan zjtes.**
- LABEL CASES BY, masukkan variabel **propinsi.**
- NUMBER OF CLUSTER atau jumlah kelompok yang akan dibentuk.

Ketik **3.**

- Klik kotak **Saves....**, kemudian aktifkan kotak **Cluster membership** dan **Distance from cluster center**, kemudian tekan CONTINUE untuk kembali ke menu utama.
- Klik kotak **Options...**,
Untuk STATISTICS, biarkan kotak INITIAL CLUSTER CENTER tetap aktif, dan aktifkan kotak **Anova Table.**

Kemudian tekan **OK** untuk proses data.



Lampiran 3: Proximity Matrix Bidang Pendidikan

Proximity Matrix

Case	Squared Euclidean Distance					Squared Euclidean Distance					11:Jawa Tengah	12:DI Yogyakarta	13:Jawa Timur
	1:Dista Aceh	2:Sumatera Utara	3:Sumatera Barat	4:Riau	5:Jambi	6:Sumatera Selatan	7:Bengkulu	8:Lampung	9:DKI Jakarta	10:Jawa Barat			
1:Dista Aceh		3.732	2.412	.413	.885	.250	.834	9.767E-02	55.233	29.641	14.672	4.745	44.667
2:Sumatera Utara	3.732		2.412	.413	.885	.250	.834	9.767E-02	55.233	29.641	14.672	4.745	44.667
3:Sumatera Barat	2.412	2.718		5.985	7.744	2.558	7.627	4.347	32.522	13.892	5.725	.773	24.682
4:Riau	.413	5.985	4.710		.162	.838	.130	.298	62.022	36.461	19.735	7.328	53.104
5:Jambi	.885	7.744	5.600	.162		1.583	9.803E-03	.690	67.490	40.619	22.713	9.159	57.570
6:Sumatera Selatan	.250	2.558	2.436	.838	1.583		1.511	.285	51.045	26.917	13.352	4.055	41.511
7:Bengkulu	.834	7.627	5.561	.130	9.803E-03	1.511		.622	66.788	40.320	22.437	9.047	57.277
8:Lampung	9.767E-02	4.347	2.901	.298	.690	.285	.622		57.319	31.578	16.131	5.782	46.831
9:DKI Jakarta	55.233	32.522	47.514	62.022	67.490	51.045	66.788	57.319		15.595	22.998	31.003	16.464
10:Jawa Barat	29.641	13.892	19.244	36.461	40.619	26.917	40.320	31.578	15.595		3.374	12.816	3.571
11:Jawa Tengah	14.672	5.725	7.651	19.735	22.713	13.352	22.437	16.131	22.998	3.374		4.415	10.008
12:DI Yogyakarta	4.745	.773	3.143	7.328	9.159	4.055	9.047	5.782	31.003	12.816	4.415		22.852
13:Jawa Timur	44.667	24.682	29.894	53.104	57.570	41.511	57.277	46.831	16.464	3.571	10.008	22.852	
14:Bali	1.091	4.256	.640	2.483	2.898	1.639	2.898	1.435	56.135	26.078	11.867	4.701	38.950
15:Nusa Tenggara Barat	.588	6.447	4.975	7.592E-02	8.067E-02	1.097	8.438E-02	.481	63.428	37.845	20.890	7.856	54.438
16:Nusa Tenggara Timur	.760	6.592	2.811	.968	.864	1.503	.872	.647	64.848	35.479	18.522	7.731	49.785
17:Kalimantan Barat	.320	5.976	3.468	.243	.283	.920	.264	.223	62.637	35.512	18.697	7.178	50.983
18:Kalimantan Tengah	.752	7.272	5.410	8.895E-02	1.868E-02	1.334	1.084E-02	.539	66.031	39.617	22.041	8.790	56.540
19:Kalimantan Selatan	.332	5.842	3.429	.258	.281	.897	.284	.255	62.652	35.402	18.744	7.062	50.713
20:Kalimantan Timur	.375	4.928	2.104	.839	1.033	.761	.999	.275	59.606	31.764	16.067	6.293	45.886
21:Sulawesi Utara	1.679	4.028	.637	3.467	4.179	2.221	4.147	2.245	53.214	22.954	9.607	3.992	36.073
22:Sulawesi Tengah	.583	6.806	4.979	7.126E-02	5.853E-02	1.227	5.552E-02	.504	64.485	38.206	20.920	8.036	55.077
23:Sulawesi Selatan	2.894	.720	.991	5.354	6.922	2.160	6.792	3.440	37.906	14.293	4.982	1.410	25.464
24:Sulawesi Tenggara	.792	7.538	5.486	.130	3.949E-02	1.494	1.329E-02	.600	65.979	39.957	22.084	8.882	56.941
25:Maluku	.352	6.027	3.545	.241	.302	.843	.289	.189	63.626	35.622	18.938	7.493	51.344
26:Irian Jaya	.717	7.005	5.382	7.099E-02	5.858E-02	1.211	3.542E-02	.473	65.023	39.150	21.734	8.585	55.922

This is a dissimilarity matrix

Lampiran 3 (Lanjutan):

Proximity Matrix

Case	Squared Euclidean Distance							Squared Euclidean Distance					
	14:Bali	15:Nusa Tenggara Barat	16:Nusa Tenggara Timur	17:Kalimantan Barat	18:Kalimantan Tengah	19:Kalimantan Selatan	20:Kalimantan Timur	21:Sulawesi Utara	22:Sulawesi Tengah	23:Sulawesi Selatan	24:Sulawesi Tenggara	25:Maluku	26:Irian Jaya
1:Dista Aceh	1.091	.588	.760	.320	.752	.332	.375	1.679	.583	2.894	.792	.352	.717
2:Sumatera Utara	4.256	6.447	6.592	5.976	7.272	5.842	4.928	4.028	6.806	.720	7.538	6.027	7.005
3:Sumatera Barat	.640	4.975	2.811	3.468	5.410	3.429	2.104	.637	4.979	.991	5.486	3.545	5.382
4:Riau	2.483	7.592E-02	.968	.243	8.895E-02	.258	.839	3.467	7.126E-02	5.354	.130	.241	7.099E-02
5:Jambi	2.898	8.067E-02	.864	.283	1.868E-02	.281	1.033	4.179	5.853E-02	6.922	3.949E-02	.302	5.858E-02
6:Sumatera Selatan	1.639	1.097	1.503	.920	1.334	.897	.761	2.221	1.227	2.160	1.494	.843	1.211
7:Bengkulu	2.898	8.438E-02	.872	.264	1.084E-02	.284	.999	4.147	5.552E-02	6.792	1.329E-02	.289	3.542E-02
8:Lampung	1.435	.481	.647	.223	.539	.255	.275	2.245	.504	3.440	.600	.189	.473
9:DKI Jakarta	56.135	63.428	64.848	62.637	66.031	62.652	59.606	53.214	64.485	37.906	65.979	63.626	65.023
10:Jawa Barat	26.078	37.845	35.479	35.512	39.617	35.402	31.764	22.954	38.206	14.293	39.957	35.622	39.150
11:Jawa Tengah	11.867	20.890	18.522	18.697	22.041	18.744	16.067	9.607	20.920	4.982	22.084	18.938	21.734
12:DI Yogyakarta	4.701	7.856	7.731	7.178	8.790	7.062	6.293	3.992	8.036	1.410	8.882	7.493	8.585
13:Jawa Timur	38.950	54.438	49.785	50.983	56.540	50.713	45.886	36.073	55.077	25.464	56.941	51.344	55.922
14:Bali		2.615	1.003	1.469	2.843	1.461	.776	.292	2.497	2.293	2.849	1.549	2.910
15:Nusa Tenggara Barat	2.615		.930	.257	5.440E-02	.233	.914	3.782	3.571E-02	5.911	9.930E-02	.291	7.324E-02
16:Nusa Tenggara Timur	1.003	.930		.252	.889	.252	.172	2.151	.874	4.923	.891	.300	.932
17:Kalimantan Barat	1.469	.257	.252		.257	1.736E-02	.260	2.510	.223	4.830	.258	4.637E-02	.272
18:Kalimantan Tengah	2.843	5.440E-02	.889	.257		.265	.950	4.077	4.952E-02	6.506	3.345E-02	.253	1.452E-02
19:Kalimantan Selatan	1.461	.233	.252	1.736E-02	.265		.273	2.552	.230	4.809	.300	5.518E-02	.284
20:Kalimantan Timur	.776	.914	.172	.260	.950	.273		1.732	.920	3.503	.994	.258	.930
21:Sulawesi Utara	.292	3.782	2.151	2.510	4.077	2.552	1.732		3.552	1.850	4.034	2.602	4.156
22:Sulawesi Tengah	2.497	3.571E-02	.874	.223	4.952E-02	.230	.920	3.552		6.036	5.320E-02	.269	8.923E-02
23:Sulawesi Selatan	2.293	5.911	4.923	4.830	6.506	4.809	3.503	1.850	6.036		6.673	4.815	6.323
24:Sulawesi Tenggara	2.849	9.930E-02	.891	.258	3.345E-02	.300	.994	4.034	5.320E-02	6.673		.313	5.687E-02
25:Maluku	1.549	.291	.300	4.637E-02	.253	5.518E-02	.258	2.602	.269	4.815	.313		.258
26:Irian Jaya	2.910	7.324E-02	.932	.272	1.452E-02	.284	.930	4.156	8.923E-02	6.323	5.687E-02	.258	

This is a dissimilarity matrix

PLAGIAT MERUPAKAN TINDAKAN TIDAK TERPUJI

Lampiran 5: Data Bidang Kesehatan

Propinsi	Dokter	SKM	APT DIII	Farmasi	Akper	Epedemiologi	Analisis Lab	Tng Elektro dan Teknik	Bidan	SPK	Asisten APT	Sanitarian
DI Aceh	567	33	51	48	280	2	235	67	283	1890	69	495
Sumatera Utara	2576	33	185	77	660	6	288	125	1075	6504	124	441
Sumatera Barat	284	61	74	13	241	9	90	60	607	1314	25	270
Riau	535	47	43	23	282	3	94	53	404	1995	33	437
Jambi	309	16	31	49	357	4	169	61	662	1254	83	388
Sumatera Selatan	883	16	28	41	432	7	160	67	294	2880	67	517
Bengkulu	178	11	22	23	180	3	67	45	136	1443	174	274
Lampung	318	70	73	5	398	3	132	77	631	1537	59	393
DKI Jakarta	4188	53	670	132	694	18	196	622	1929	5414	1258	396
Jawa Barat	3342	70	541	103	1019	8	258	345	1825	9297	1055	435
Jawa Tengah	3002	96	344	153	1560	7	334	264	1617	8254	977	481
DI Yogyakarta	1225	40	267	51	319	9	251	228	295	1857	987	339
Jawa Timur	2795	187	585	66	1049	6	283	521	3398	5274	1282	136
Kkalimantan Barat	580	14	28	81	170	1	113	106	173	1447	96	419
Kalimantan Tengah	332	10	16	19	164	5	70	76	155	1577	52	352
Kalimantan Selatan	498	13	19	54	145	1	126	58	219	1942	73	438
Kalimantan Timur	239	12	16	57	102	4	56	62	147	1629	77	387
Sulawesi Utara	334	15	12	43	193	1	36	46	126	1675	85	433
Sulawesi Tengah	387	19	25	29	141	2	54	34	111	3380	47	212
Sulawesi Selatan	1400	44	35	63	462	3	241	38	698	1377	178	379
Sulawesi Tenggara	300	24	11	27	149	2	85	32	109	1637	66	256
Bali	690	50	218	40	359	2	100	63	711	1818	467	592
Nusa Tenggara Barat	414	71	31	20	124	1	128	52	231	1416	74	361
Nusa Tenggara Timur	297	11	98	27	172	2	58	84	116	1822	44	398
Maluku	406	9	31	28	219	4	57	45	117	1960	89	111
Irian Jaya	838	17	12	42	462	2	155	21	34	264	46	124
Indonesia	26917	1042	3466	1314	10333	115	3836	3252	16103	70857	7587	9464

Sumber: Biro Kepegawaian DepKes RI

Lampiran 6: Proximity Matrix Bidang_Kesehatan

Proximity Matrix

Case	Squared Euclidean Distance					Squared Euclidean Distance					Squared Euclidean Distance				
	1:DI Aceh	2:Sumatera Utara	3:Sumatera Barat	4:Riau	5:Jambi	6:Sumatera Selatan	7:Bengkulu	8:Lampung	9:DKI Jakarta	10:Jawa Barat	11:Jawa Tengah	12:DI Yogyakarta	13:Jawa Timur	14:Kalimantan Barat	15:Kalimantan Tengah
1:DI Aceh		14.612	11.644	3.621	2.311	3.378	8.383	5.007	76.399	36.291	56.135	12.738	79.342	3.818	6.937
2:Sumatera Utara	14.612		25.980	20.017	17.088	12.192	28.000	21.617	42.094	18.305	21.591	16.637	51.631	20.720	25.298
3:Sumatera Barat	11.644	25.980		4.996	6.160	8.619	5.035	4.196	72.819	43.112	72.463	13.653	72.511	11.959	3.953
4:Riau	3.621	20.017	4.996		2.547	3.620	3.302	1.286	79.171	41.318	65.536	15.395	79.074	4.206	2.136
5:Jambi	2.311	17.088	6.160	2.547		3.143	3.694	3.866	73.267	36.351	60.146	11.554	77.525	2.786	2.987
6:Sumatera Selatan	3.378	12.192	8.619	3.620	3.143		8.246	6.331	65.090	34.177	52.799	11.960	81.207	6.146	4.999
7:Bengkulu	8.383	28.000	5.035	3.302	3.694	8.246		5.133	86.898	50.377	79.507	16.464	89.004	5.020	.870
8:Lampung	5.007	21.617	4.196	1.286	3.866	6.331	5.133		79.799	39.556	64.830	14.729	69.271	8.033	4.407
9:DKI Jakarta	76.399	42.094	72.819	79.171	73.267	65.090	86.898	79.799		15.794	31.949	39.319	38.255	77.035	81.732
10:Jawa Barat	36.291	18.305	43.112	41.318	36.351	34.177	50.377	39.556	15.794		15.450	17.864	24.267	40.649	48.356
11:Jawa Tengah	56.135	21.591	72.463	65.536	60.146	52.799	79.507	64.830	31.949	15.450		43.222	35.078	63.650	77.448
12:DI Yogyakarta	12.738	16.637	13.653	15.395	11.554	11.960	16.464	14.729	39.319	17.864	43.222		50.741	15.984	15.446
13:Jawa Timur	79.342	51.631	72.511	79.074	77.525	81.207	89.004	69.271	38.255	24.267	35.078	50.741		86.682	90.363
14:Kalimantan Barat	3.818	20.720	11.959	4.206	2.786	6.146	5.020	8.033	77.035	40.649	63.650	15.984	86.682		4.888
15:Kalimantan Tengah	6.937	25.298	3.953	2.136	2.987	4.999	.870	4.407	81.732	48.356	77.448	15.446	90.363	4.888	
16:Kalimantan Selatan	2.364	19.753	10.155	2.208	1.952	4.373	3.595	5.512	81.890	43.442	66.494	16.134	88.581	.820	3.141
17:Kalimantan Timur	6.135	24.381	6.689	2.700	2.783	5.174	2.013	6.482	79.620	46.573	72.746	16.324	91.817	1.839	1.388
18:Sulawesi Utara	6.057	26.460	9.509	2.041	3.929	6.318	2.554	5.729	88.294	49.611	75.413	20.395	95.017	2.188	2.283
19:Sulawesi Tengah	11.192	25.383	7.148	5.205	6.732	10.828	1.668	7.456	87.382	52.742	75.962	20.193	85.967	7.005	3.189
20:Sulawesi Selatan	2.539	11.695	10.197	5.619	2.548	5.484	9.132	5.877	66.165	27.435	46.839	10.062	64.511	5.358	8.826
21:Sulawesi Tenggara	7.658	26.851	5.269	3.110	3.794	8.886	.372	4.612	89.238	50.927	78.313	17.471	87.100	4.716	1.580
22:Bali	5.324	19.692	13.270	4.080	6.591	6.002	10.962	5.829	68.120	30.723	54.698	14.364	72.079	7.003	9.146
23:Nusa Tenggara Barat	4.728	24.819	5.884	1.608	4.402	8.369	3.876	1.510	87.738	45.971	72.003	16.487	77.440	5.592	4.177
24:Nusa Tenggara Timur	5.761	24.854	7.164	1.646	3.353	5.670	1.516	4.434	83.676	47.282	75.275	17.320	90.246	3.115	1.067
25:maluku	15.403	30.841	6.279	8.706	8.098	14.351	2.096	10.030	87.028	53.934	82.296	19.575	86.652	10.084	4.254
26:Irian Jaya	11.787	28.631	9.183	9.658	6.437	14.560	4.489	9.604	89.302	48.154	75.563	18.058	83.192	8.997	7.202

This is a dissimilarity matrix

Lampiran 6 (Lanjutan):

Proximity Matrix

Case	Squared Euclidean Distance					Squared Euclidean Distance				Squared Euclidean	
	16:Kalimantan Selatan	17:Kalimantan Timur	18:Sulawesi Utara	19:Sulawesi Tengah	20:Sulawesi Selatan	21:Sulawesi Tenggara	22:Bali	23:Nusa Tenggara Barat	24:Nusa Tenggara Timur	25:maluku	26:Irian Jaya
1:DI Aceh	2.364	6.135	6.057	11.192	2.539	7.658	5.324	4.728	5.761	15.403	11.787
2:Sumatera Utara	19.753	24.381	26.460	25.383	11.695	26.851	19.692	24.819	24.854	30.841	28.631
3:Sumatera Barat	10.155	6.689	9.509	7.148	10.197	5.269	13.270	5.884	7.164	6.279	9.183
4:Riau	2.208	2.700	2.041	5.205	5.619	3.110	4.080	1.608	1.646	8.706	9.658
5:Jambi	1.952	2.783	3.929	6.732	2.548	3.794	6.591	4.402	3.353	8.098	6.437
6:Sumatera Selatan	4.373	5.174	6.318	10.828	5.484	8.886	6.002	8.369	5.670	14.351	14.560
7:Bengkulu	3.595	2.013	2.554	1.668	9.132	.372	10.962	3.876	1.516	2.096	4.489
8:Lampung	5.512	6.482	5.729	7.456	5.877	4.612	5.829	1.510	4.434	10.030	9.604
9:DKI Jakarta	81.890	79.620	88.294	87.382	66.165	89.238	68.120	87.738	83.676	87.028	89.302
10:Jawa Barat	43.442	46.573	49.611	52.742	27.435	50.927	30.723	45.971	47.282	53.934	48.154
11:Jawa Tengah	66.494	72.746	75.413	75.962	46.839	78.313	54.698	72.003	75.275	82.296	75.563
12:DI Yogyakarta	16.134	16.324	20.395	20.193	10.062	17.471	14.364	16.487	17.320	19.575	18.058
13:Jawa Timur	88.581	91.817	95.017	85.967	64.511	87.100	72.079	77.440	90.246	86.652	83.192
14:Kalimantan Barat	.820	1.839	2.188	7.005	5.358	4.716	7.003	5.592	3.115	10.084	8.997
15:Kalimantan Tengah	3.141	1.388	2.283	3.189	8.826	1.580	9.146	4.177	1.067	4.254	7.202
16:Kalimantan Selatan		1.588	1.261	5.433	5.014	3.329	5.615	3.690	1.642	9.210	8.904
17:Kalimantan Timur	1.588		1.095	4.005	8.072	2.450	7.813	4.882	1.279	6.065	8.503
18:Sulawesi Utara	1.261	1.095		4.509	11.454	2.817	6.050	4.108	.717	8.030	9.902
19:Sulawesi Tengah	5.433	4.005	4.509		11.454	1.181	14.687	5.373	3.384	1.708	5.889
20:Sulawesi Selatan	5.014	8.072	9.068	11.454		8.071	8.198	6.102	8.870	13.234	7.791
21:Sulawesi Tenggara	3.329	2.450	2.817	1.181	8.071		11.519	2.680	1.925	2.080	3.656
22:Bali	5.615	7.813	6.050	14.687	8.198	11.519		7.197	6.371	20.147	19.873
23:Nusa Tenggara Barat	3.690	4.882	4.108	5.373	6.102	2.680	7.197		3.550	8.409	7.945
24:Nusa Tenggara Timur	1.642	1.279	.717	3.384	8.870	1.925	6.371	3.550		6.145	8.593
25:maluku	9.210	6.065	8.030	1.708	13.234	2.080	20.147	8.409	6.145		3.305
26:Irian Jaya	8.904	8.503	9.902	5.889	7.791	3.656	19.873	7.945	8.593	3.305	

This is a dissimilarity matrix

Lampiran 7: Dendrogram Bidang_Kesehatan

***** HIERARCHICAL CLUSTER ANALYSIS *****

Dendrogram using Average Linkage (Between Groups)

