

# TEORI DAN LANGKAH-LANGKAH ANALISIS GEROMBOL MENGUNAKAN APLIKASI MINITAB DAN SPSS ANALISIS PEUBAH GANDA

---

**Audhi Aprilliant (G14160021)**

Departemen Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam,  
Institut Pertanian Bogor, Jl. Meranti Wing 22 Lantai 4,  
Dramaga, Bogor, Jawa Barat 16680

## TINJAUAN PUSTAKA

### Definisi Analisis Cluster

Menurut Cornish di dalam Ariska N (2017), analisis gerombol merupakan suatu metode multivariat yang bertujuan untuk mengelompokkan sampel subjek atas dasar satu set peubah yang diukur menjadi beberapa kelompok yang berbeda sehingga subjek yang sama ditempatkan dalam kelompok yang sama. Menurut Han & Kamber (2001), analisis gerombol adalah suatu teknik *data mining* untuk mengelompokkan himpunan objek (dataset) ke dalam beberapa gerombol hanya berdasarkan kemiripan karakteristik dari atribut yang dimiliki oleh data objek sedemikian sehingga data objek yang berada di dalam gerombol yang sama memiliki kemiripan satu sama lain, tetapi tidak mirip dengan data objek yang berada dalam gerombol yang berbeda.

### Ukuran Kemiripan

Ukuran kemiripan digunakan untuk mencari pasangan objek yang mirip dalam data. Kemiripan antar pasangan objek  $X$  dan  $Y$  dinyatakan dengan  $si(X, Y)$  yang akan bernilai besar apabila  $X$  dan  $Y$  merupakan pasangan objek yang mirip. Sebaliknya  $si(X, Y)$  akan bernilai kecil apabila  $X$  dan  $Y$  merupakan pasangan objek yang tidak mirip. Untuk setiap pasangan objek  $X$  dan  $Y$  berlaku tiga kondisi sebagai berikut:

- 1)  $0 \leq si(X, Y) \leq 1$ , kemiripan bernilai 0 dan 1
- 2)  $si(X, Y) = 1$ , setiap objek mirip dengan dirinya sendiri
- 3)  $si(X, Y) = si(Y, X)$ , kemiripan bersifat simetri

### Ukuran Ketidakmiripan

Ukuran ketidakmiripan digunakan untuk mencari jarak antara pasangan objek di dalam data. Jarak antara pasangan objek  $X$  dan  $Y$  dinyatakan dengan  $d(X, Y)$  yang akan bernilai besar apabila  $X$  dan  $Y$  merupakan pasangan objek yang tidak mirip. Sebaliknya  $d(X, Y)$  akan bernilai kecil apabila  $X$  dan  $Y$  merupakan pasangan objek yang mirip. Untuk setiap pasangan objek  $X$  dan  $Y$  berlaku tiga kondisi sebagai berikut:

- 1)  $d(X, Y) \geq 0$ , jarak merupakan bilangan non-negatif
- 2)  $d(X, Y) = 0$ , jarak suatu objek dengan dirinya sendiri
- 3)  $d(X, Y) = d(Y, X)$ , jarak bersifat simetri

### Metode Penggerombolan Data Numerik Hierarki

Penggerombolan data numerik dilakukan berdasarkan ukuran ketidakmiripan atau jarak untuk data numerik. Hasil penggerombolan disajikan dalam bentuk dendrogram (diagram pohon) yang memungkinkan penelusuran objek-objek yang diamati menjadi lebih mudah dan informatif maupun plot gerombol melalui *Principal Component Analysis*. Pada peubah yang memiliki jenis skala data numerik maka jarak yang dapat digunakan adalah jarak euclidean.

Jarak euclidean digunakan dalam mengukur jumlah kuadrat perbedaan nilai pada masing-masing peubah.

$$d(X, Y) = \sqrt{\sum_{i=1}^n (X_i - Y_i)^2}$$

Keterangan:

$d(X, Y)$  : Jarak antara objek  $X$  dengan  $Y$

$n$  : Jumlah observasi atau amatan

$X_i$  : Data objek  $X$  ke- $i$

$Y_i$  : Data objek  $Y$  ke- $i$

### Metode Penggerombolan Data Numerik *K-Means*

Menurut Purnamaningsih C (2014), *K-Means* merupakan salah satu metode data penggerombolan non-hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih gerombol atau kelompok. Metode ini mempartisi data ke dalam gerombol atau kelompok sehingga data yang memiliki karakteristik sama dikelompokkan ke dalam satu gerombol yang sama. Menurut MacQueen JB (1967), berikut adalah langkah-langkah dari algoritma *K-Means*:

- 1) Menentukan banyak gerombol yang ingin dibentuk
- 2) Membangkitkan nilai acak untuk pusat k-gerombol awal (*centorid*) sebanyak k-gerombol
- 3) Menghitung jarak setiap data input terhadap masing-masing centorid menggunakan rumus jarak (*Euclidian Distance*) hingga ditemukan jarak yang paling dekat dari setiap data dengan *centorid*
- 4) Mengklasifikasikan setiap data berdasarkan kedekatannya dengan *centroid* (jarak terdekat)
- 5) Memperbaharui nilai *centroid*. Nilai *centorid* baru diperoleh dari rata-rata gerombol yang bersangkutan dengan menggunakan rumus:

$$C_k = \frac{1}{n} \sum_{i=1}^n d_i$$

Keterangan:

$k$  : Banyaknya gerombol dimana bernilai  $1, 2, \dots, k$

$n$  : Jumlah data dalam gerombol

$d_i$  : Jumlah dari nilai jarak yang masuk dalam masing-masing gerombol

- 6) Melakukan perulangan dari langkah 2) sampai 5) hingga anggota setiap gerombol tidak ada yang berubah
- 7) Apabila di langkah 6) telah terpenuhi, maka nilai rata-rata pusat gerombol pada iterasi terakhir akan digunakan sebagai parameter untuk menentukan klasifikasi data

## CONTOH PENERAPAN ANALISIS GEROMBOL MENGUNAKAN SPSS

Dari penelitian yang dilakukan terhadap 12 kota, ingin diketahui pengelompokan kota-kota tersebut berdasarkan instrumen 5 peubah, yaitu:

- a. Jumlah pendapatan kota (trilyun)
- b. Jumlah pinjaman pemerintah kota (milyar)
- c. Jumlah dana hibah yang dimiliki kota (milyar)
- d. Jumlah konsumsi pemerintah kota (milyar)
- e. Jumlah penduduk kota (juta jiwa)

Untuk itu data yang berhasil dikumpulkan sebagai berikut:

Kota	Pendapatan	Pinjaman	Dana Hibah	Konsumsi	Produk
A	55	5,6	9	50	25
B	61	8	7	62	41
C	58	3,9	7	60	32
D	67	5,5	7	64	51
E	71	5,7	6	70	42
F	76	7,6	8	80	29
G	81	8,7	9	80	57
H	56	7,1	6	86	29
I	84	7,6	7	82	46
J	88	6,5	8	86	52
K	84	6,8	9	88	61
L	90	8	9	90	66

### 1. Melakukan standarisasi atau transformasi

Mengingat data yang terkumpul mempunyai variabilitas satuan, maka perlu dilakukan langkah standarisasi atau transformasi terhadap peubah yang relevan ke bentuk Z-Score, sebagai berikut:

- a. Setelah keseluruhan data yang dikumpulkan tersebut diatas di-entry dalam program SPSS, selanjutnya klik menu **analyze** dan pilih sub menu **Descriptives Statistics** lalu **Descriptives**.
- b. Masukkan ke dalam kotak **Variable(s)** seluruh peubah instrumen penilai, yaitu peubah jumlah pendapatan, jumlah pinjaman, jumlah dana hibah, jumlah konsumsi, dan jumlah penduduk (dalam hal ini peubah kota tidak dimasukkan karena data bertipe string). Kemudian aktifkan bagian **Save standardized values as variables**. Abaikan bagian yang lain lalu tekan **OK**.

**Descriptive Statistics**

	N	Minimum	Maximum	Mean	Std. Deviation
Pendapatan	12	55.00	90.00	72.5833	12.96470
Pinjaman	12	4	9	6.75	1.371
Dana_Hibah	12	6.00	9.00	7.6667	1.15470
Konsumsi	12	50.00	90.00	74.8333	13.14142
Produk	12	25.00	66.00	44.2500	13.57220
Valid N (listwise)	12				

Kota	Pendapatan	Pinjaman	Dana Hibah	Konsumsi	Produk
A	-1,35625	-0,83893	1,154701	-1,8897	-1,41834
B	-0,89345	0,911877	-0,57735	-0,97656	-0,23946
C	-1,12485	-2,07908	-0,57735	-1,12875	-0,90258
D	-0,43066	-0,91188	-0,57735	-0,82437	0,49734
E	-0,12213	-0,76598	-1,44338	-0,36779	-0,16578
F	0,263536	0,620076	0,288675	0,393159	-1,12362
G	0,649199	1,422528	1,154701	0,393159	0,93942
H	-1,27911	0,255325	-1,44338	0,84973	-1,12362
I	0,880597	0,620076	-0,57735	0,545349	0,12894
J	1,189127	-0,18238	0,288675	0,84973	0,57102
K	0,880597	0,036475	1,154701	1,001921	1,234141
L	1,343392	0,911877	1,154701	1,154112	1,602541

Catatan:

Untuk selanjutnya, hasil z-score inilah yang akan dipakai dasar analisis cluster. Namun apabila data yang terkumpul tidak mempunyai variabilitas satuan, maka prose analisis cluster dapat langsung dilakukan tanpa terlebih dahulu melakukan transformasi atau standarisasi.

## 2. Metode K-Means Cluster (Non-Hierarki)

Sebagaimana telah dijelaskan sebelumnya bahwa metode K-Means Cluster ini jumlah cluster ditentukan sendiri. Perlu diingat bahwa bahan analisis bukan lagi data asli, namun data hasil transformasi/ standarisasi.

- Buka menu **Analyze**, lalu pilih sub menu **Classify** dan pilih **K-Means Cluster**
- Masukkan seluruh peubah Z-Score ke dalam kotak **Variable(s)**. Kemudian peubah Kota dimasukkan dalam kotak **Label Cases by**. **Number of Clusters** dalam hal ini diisi menurut jumlah cluster yang akan dibentuk dalam penelitian yang dimaksud. Dalam hal ini diisi **3**, berarti diharapkan akan dibentuknya **3 cluster**. Kemudian klik pada kotak **Save**.

Catatan:

Kotak dialog Save memungkinkan hasil cluster disimpan dalam bentuk peubah baru. Hal ini berguna untuk proses profiling cluster, yang akan dilakukan pada tahapan akhir analisis cluster.

- Aktifkan kedua kotak dalam menu **Save**, yaitu **Cluster membership** dan **Distance from cluster center**. Selanjutnya pilih **Continue** untuk kembali ke menu utama.
- Kemudian pilih pada kotak **Options**. Pada bagian **Statistics**, aktifkan **Initial cluster centers** dan **ANOVA table**. Abaikan bagian yang lain, lalu tekan **Continue** untuk kembali ke menu utama. Dari tampilan menu utama cluster, abaikan bagian yang lain lalu tekan tombol **OK**.

Initial Cluster Centers

	Cluster		
	1	2	3
Zscore(Pendapatan)	1.34339	-1.27911	-1.12485
Zscore(Pinjaman)	.91188	.25533	-2.07908
Zscore(Dana_Hibah)	1.15470	-1.44338	-.57735
Zscore(Konsumsi)	1.15411	.84973	-1.12875
Zscore(Produk)	1.60254	-1.12362	-.90258

Tabel di atas merupakan tampilan pertama proses clustering data sebelum dilakukan iterasi. Untuk mendeteksi berapa kali proses iterasi yang dilakukan dalam proses clustering dari 12 obyek yang diteliti, dapat dilihat dari tampilan output berikut ini:

**Iteration History<sup>a</sup>**

Iteration	Change in Cluster Centers		
	1	2	3
1	1.073	1.395	1.103
2	.000	.000	.000

a. Convergence achieved due to no or small change in cluster centers. The maximum absolute coordinate change for any center is .000. The current iteration is 2. The minimum distance between initial centers is 3,192.

Proses clustering yang dilakukan melalui 2 tahapan iterasi untuk mendapatkan cluster yang tepat. Dari tabel di atas disebutkan bahwa jarak minimum antar pusat cluster yang terjadi dari hasil iterasi adalah 3.192. Adapun hasil akhir dari proses clustering digambarkan berikut ini:

**Final Cluster Centers**

	Cluster		
	1	2	3
Zscore(Pendapatan)	.98858	-.63634	-.75847
Zscore(Pinjaman)	.56172	.59576	-1.14896
Zscore(Dana_Hibah)	.63509	-.57735	-.36084
Zscore(Konsumsi)	.78885	.08878	-1.05265
Zscore(Produk)	.89521	-.82890	-.49734

Output Final Cluster Centers tersebut di atas masih terkait dengan proses standarisasi data sebelumnya, yang mengacu pada z-score dengan ketentuan sebagai berikut:

- Nilai negatif (-) apabila data berada di bawah rata-rata total
- Nilai positif (+) apabila data berada di atas rata-rata total

Rumus umum yang digunakan yaitu:

$$X = \mu + Z\sigma$$

Keterangan:

$X$  : Rata – rata sampel (peubah dalam cluster)

$\mu$  : Rata – rata populasi

$Z$  : Nilai standarisasi

$\sigma$  : Standar deviasi

Contoh:

Apabila ingin diketahui rata-rata jumlah pendapatan kota di cluster-1 yaitu:

$$\begin{aligned}
 &= (\text{Rata – rata pendapatan seluruh kota}) + \\
 &\quad (0.98858 \times \text{standar deviasi rata – rata pendapatan}) \\
 &= 72.58 + (0.98858 \times 12.965) \\
 &= 85.3969
 \end{aligned}$$

Jadi, rata-rata jumlah pendapatan kota yang berada di cluster-1 adalah Rp85.3969 trilyun.

Dari tabel output Final Cluster Centers, dengan ketentuan yang telah dijabarkan di atas pula, dapat didefinisikan sebagai berikut:

a. Cluster-1

Dalam cluster-1 ini berisikan kota-kota yang mempunyai jumlah pendapatan kota, jumlah pinjaman, jumlah dana hibah, jumlah konsumsi, dan jumlah penduduk yang lebih dari rata-rata populasi kota yang diteliti. Hal ini terbukti dari nilai positif (+) yang terdapat pada tabel Final Cluster Centers dalam keseluruhan variabel. Dengan demikian, dapat diduga bahwa cluster-1 ini merupakan pengelompokan dari kota-kota besar.

b. Cluster-2

Karakteristik kota yang masuk dalam pengelompokan cluster-2 yaitu memiliki rata-rata jumlah pinjaman dan jumlah konsumsi yang melebihi rata-rata populasi kota yang diteliti. Untuk instrumen peubah yang lain kota-kota di cluster-2 ini berada di atas rata-rata populasi. Dengan demikian, dapat diduga sekumpulan kota-kota menengah berada pada cluster-2.

c. Cluster-3

Karakteristik kota-kota yang mengelompok pada cluster-3 adalah keseluruhan instrumen penilai berada pada posisi dibawah rata-rata populasi kota yang diteliti. Sehingga dapat diduga bahwa cluster-3 merupakan pengelompokan kota-kota kecil.

Catatan:

**Penamaan cluster atau penarikan kesimpulan sangat bersifat subyektif dan bergantung pada tujuan penelitian.**

Tahapan selanjutnya yang perlu dilakukan yaitu melihat perbedaan peubah pada cluster yang terbentuk. Dalam hal ini dapat dilihat dari **nilai F** dan **nilai probabilitas (sig)** masing-masing peubah, seperti tampak dalam tabel berikut:

**ANOVA**

	Cluster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
Zscore(Pendapatan)	4.201	2	.289	9	14.556	.002
Zscore(Pinjaman)	3.961	2	.342	9	11.587	.003
Zscore(Dana_Hibah)	1.769	2	.829	9	2.133	.174
Zscore(Konsumsi)	3.784	2	.381	9	9.921	.005
Zscore(Produk)	3.529	2	.438	9	8.056	.010

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.

Rumus nilai F:

$$F = \frac{\text{MS Between}}{\text{MS Within}}$$

Catatan:

Dimana dalam tabel ANOVA di atas **MS Between** ditunjukkan oleh **Means Square dalam kolom Cluster**, sedangkan **MS Within** ditunjukkan oleh **Means Square dalam kolom Error**.

Semakin besar nilai F dan ( $\text{sig} < 0.05$ ), maka semakin besar perbedaan peubah pada cluster yang terbentuk.

Dengan demikian hasil cluster yang didapat dalam penelitian ini bahwa untuk instrumen jumlah pendapatan kota yang paling menunjukkan adanya perbedaan di antara kota-kota pada ketiga cluster yang terbentuk. Hal ini dengan ditunjukkannya nilai  $F = 14.556$  dan  $\text{sig} = 0.002$ . Dan untuk peubah yang lain pun dapat didefinisikan lebih lanjut.

Selanjutnya untuk mengetahui jumlah anggota masing-masing cluster yang terbentuk dapat dilihat pada tabel output berikut ini:

**Number of Cases in each Cluster**

Cluster	1	5.000
	2	3.000
	3	4.000
Valid		12.000
Missing		.000

### 3. Metode Hierarchical Cluster (Hierarki)

Konsep dari metode hirarkis ini dimulai dengan menggabungkan 2 obyek yang paling mirip, kemudian gabungan 2 obyek tersebut akan bergabung lagi dengan satu atau lebih obyek yang paling mirip lainnya. Proses clustering ini pada akhirnya akan **menggumpal** menjadi satu cluster besar yang mencakup semua obyek. Metode ini disebut juga sebagai metode **agglomerative** yang digambarkan dengan dendogram.

- Buka menu **Analyze** lalu pilih sub menu **Classify** kemudian **Hierarchical Cluster**
- Masukkan seluruh peubah yang telah distandarisasi (Z-score) ke dalam bagian **Variable(s)**. Untuk bagian **Label Cases by** isi dengan peubah kota. Sedangkan untuk bagian **Cluster** pilih **Cases**. Pada bagian **Display** pilih keduanya yaitu **Statistics** dan **Plots**.
- Kemudian pilih pada kotak **Statistics**. Selain kotak **Agglomeration Schedule**, aktifkan pula kotak **Proximity matrix** untuk menampilkan jarak antar peubah. Pada bagian **Cluster Membership**, pilih pada pilihan **Range of Solutions** lalu ketik **2** pada **FROM** dan **4** pada **THROUGH** (Akan ditampilkan susunan 2, 3, dan 4 cluster). Kemudian tekan tombol **Continue** untuk kembali ke menu utama.
- Selanjutnya pilih pada kotak **Plots**. Aktifkan pilihan **Dendogram**. Kemudian pada bagian **Icele** pilih **None**. Abaikan bagian yang lain lalu tekan tombol **Continue** untuk kembali ke menu utama.
- Berikutnya pilih pada kotak **Method**. Pada bagian **Cluster Method** pilih **Between groups linkage**. Kemudian buka kotak **combo Square Euclidean distance** pada **Measure**, dan pada **Transform Values** buka kotak combo pada pilihan Z-score. Abaikan bagian yang lain lalu tekan tombol **Continue** untuk kembali ke menu utama. Dari tampilan menu utama, tekan tombol **OK**.

**Case Processing Summary<sup>a</sup>**

Cases					
Valid		Missing		Total	
N	Percent	N	Percent	N	Percent
12	100.0%	0	.0%	12	100.0%

a. Squared Euclidean Distance used

Tabel output di atas menunjukkan bahwa semua data sejumlah 12 obyek telah diproses tanpa ada data yang hilang.

Proximity Matrix												
Case	Squared Euclidean Distance											
	1:A	2:B	3:C	4:D	5:E	6:F	7:G	8:H	9:I	10:J	11:K	12:L
1:A	.000	8.503	5.437	8.667	12.163	10.801	19.906	15.545	18.456	19.122	21.167	28.744
2:B	8.503	.000	9.462	4.106	4.536	4.832	8.906	5.447	5.684	10.277	12.999	15.936
3:C	5.437	9.462	.000	3.897	4.602	12.328	24.118	10.186	15.174	15.788	20.603	29.525
4:D	8.667	4.106	3.897	.000	1.515	7.689	11.293	8.262	6.078	6.714	9.497	14.609
5:E	12.163	4.536	4.602	1.515	.000	6.566	13.935	4.782	4.597	7.085	12.235	17.156
6:F	10.801	4.832	12.328	7.689	6.566	.000	5.799	5.721	2.723	4.581	7.401	10.012
7:G	19.906	8.906	24.118	11.293	13.935	5.799	.000	16.295	4.378	3.961	2.432	1.761
8:H	15.545	5.447	10.186	8.262	4.782	5.721	16.295	.000	7.209	12.156	17.044	21.583
9:I	18.456	5.684	15.174	6.078	4.597	2.723	4.378	7.209	.000	1.777	4.771	5.841
10:J	19.122	10.277	15.788	6.714	7.085	4.581	3.961	12.156	1.777	.000	1.356	3.128
11:K	21.167	12.999	20.603	9.497	12.235	7.401	2.432	17.044	4.771	1.356	.000	1.139
12:L	28.744	15.936	29.525	14.609	17.156	10.012	1.761	21.583	5.841	3.128	1.139	.000

This is a dissimilarity matrix

Tabel di atas menunjukkan matrik jarak antara peubah satu dengan peubah yang lain. Semakin kecil jarak euclidean, maka semakin mirip kedua peubah tersebut sehingga akan membentuk kelompok (cluster).

Agglomeration Schedule						
Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	11	12	1.139	0	0	4
2	4	5	1.515	0	0	6
3	9	10	1.777	0	0	5
4	7	11	2.097	0	1	7
5	6	9	3.652	0	3	7
6	3	4	4.249	0	2	9
7	6	7	5.183	5	4	11
8	2	8	5.447	0	0	9
9	2	3	6.889	8	6	10
10	1	2	10.063	0	9	11
11	1	6	13.662	10	7	0

Tabel di atas merupakan hasil proses clustering dengan metode Between Group Linkage. Setelah jarak antar peubah diukur dengan jarak euclidean, maka dilakukan pengelompokan yang dilakukan secara bertingkat.

- Stage 1: terbentuk 1 cluster yang beranggotakan Kota K dan Kota L dengan jarak 1.139 (perhatikan pada kolom Coefficients). Karena proses aglomerasi dimulai dari 2 obyek yang terdekat, maka jarak tersebut adalah yang terdekat dari sekian kombinasi jarak 12 obyek yang ada. Selanjutnya lihat kolom terakhir (Next Stage), terlihat angka 4. Hal ini berarti clustering selanjutnya dilakukan dengan melihat stage 4, dengan penjelasan berikut.
- Baris ke-4 (stage 4): terlihat obyek ke-7 (Kota G) membentuk cluster dengan Kota K. Dengan demikian, sekarang cluster terdiri dari 3 obyek yaitu Kota G, K, dan L. Sedangkan jarak sebesar 2.097 merupakan jarak rata-rata obyek terakhir yang bergabung dengan 2 obyek sebelumnya, seperti tampak dalam Proximity matrix dan dapat dihitung sebagai berikut :
  - Jarak kota G dan K = 2.432
  - Jarak kota G dan L = 1.761



- c) Jarak rata-rata =  $(2.432 + 1.761)/2 = 2.0965$
- c. Stage 2: terjadi pembentukan cluster Kota D dan Kota E (berjarak 1.515) yang kemudian berlanjut ke stage 6.
- d. Demikian seterusnya dari stage 3 dilanjutkan ke stage 5, sampai ke stage terakhir.

Proses aglomerasi ini bersifat kompleks, khususnya perhitungan koefisien yang melibatkan sekian banyak obyek dan terus bertambah. Proses aglomerasi pada akhirnya akan menyatukan semua obyek menjadi satu cluster. Hanya saja dalam prosesnya dihasilkan beberapa cluster dengan masing-masing anggotanya, tergantung jumlah cluster yang dibentuk. Perincian jumlah cluster dengan anggota yang terbentuk dapat dilihat pada tabel output berikut ini:

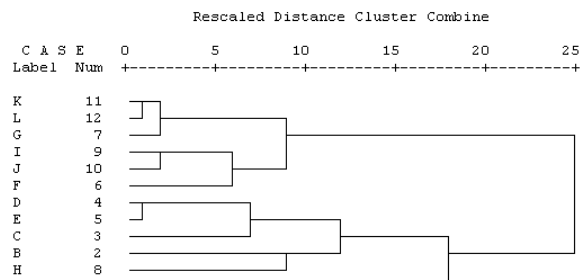
Cluster Membership			
Case	4 Clusters	3 Clusters	2 Clusters
1:A	1	1	1
2:B	2	2	1
3:C	3	2	1
4:D	3	2	1
5:E	3	2	1
6:F	4	3	2
7:G	4	3	2
8:H	2	2	1
9:I	4	3	2
10:J	4	3	2
11:K	4	3	2
12:L	4	3	2

Dari tabel di atas dapat dijabarkan bahwa:

- a. Apabila diinginkan dibentuk 4 cluster, maka:
  - a) Anggota cluster 1 adalah Kota A
  - b) Anggota cluster 2 adalah Kota B dan Kota H
  - c) Anggota cluster 3 adalah C, D, dan E
  - d) Anggota cluster 4 adalah Kota F, G, I, J, K, dan L
- b. Apabila ditentukan dibentuk 3 cluster, maka:
  - a) Anggota cluster 1 adalah Kota A
  - b) Anggota cluster 2 adalah Kota B, C, D, E, dan H
  - c) Anggota cluster 3 adalah Kota F, G, I, J, K, dan L
- c. Apabila ditentukan dibentuk 2 cluster, maka:
  - a) Anggota cluster 1 adalah Kota A, B, C, D, E, dan H
  - b) Anggota cluster 2 adalah Kota F, G, I, J, K, dan L

\*\*\*\*\* H I E R A R C H I C A L C L U S T E R A N A L Y S I S \*\*\*\*\*  
 \*

Dendrogram using Average Linkage (Between Groups)



Dendrogram berguna untuk menunjukkan anggota cluster yang ada jika akan ditentukan berapa cluster yang seharusnya dibentuk. Sebagai contoh yang terlihat dalam dendrogram, apabila akan dibentuk 2 cluster, maka cluster 1 beranggotakan Kota K sampai dengan Kota F (sesuai urutan dalam dendrogram) dan cluster2 beranggotakan Kota D sampai dengan Kota A. Demikian seterusnya dapat dengan mudah dilihat anggota tiap cluster sesuai jumlah cluster yang diinginkan.

## DAFTAR PUSTAKA

- Ariska N. 2017. *Analisis Cluster dengan Metode Ensemble Rock untuk Data berskala Campuran Kategorik dan Numerik (Kasus: Mahasiswa Aktif Program Studi Statistika FMIPA UNM)* [skripsi]. Makassar (ID): Program Studi Matematika Fakultas Matematika dan IPA Universitas Negeri Makassar.
- Safitri D, Widiharih T, Wilandari Y, Saputra AH. 2012. Analisis cluster pada kabupaten/ kota di Jawa Tengah berdasarkan Provinsi Palawija. *Media Statistika*. 5(1): 11– 16.
- Sunarso YE. 2008. *Analisis Cluster dan Aplikasinya* [skripsi]. Yogyakarta (ID): Jurusan Matematika Fakultas Sains dan Teknologi Universitas Sanata Dharma.