

Tujuan Praktikum

Pada modul 6 ini, tujuan yang hendak dicapai dalam pelaksanaan praktikum antara lain :

- Mahasiswa mampu mengenali karakteristik analisis cluster.
- Mahasiswa memahami kegunaan analisis cluster.
- Mahasiswa mengetahui asumsi dari analisis cluster dan mampu menginterpretasikan output dari model analisis cluster

Materi Praktikum

Tujuan dari **Analisis Cluster** adalah *mengelompokkan obyek berdasarkan kesamaan karakteristik* di antara obyek-obyek tersebut. Dengan demikian, ciri-ciri suatu cluster yang baik yaitu mempunyai :

- **Homogenitas internal** (within cluster); yaitu kesamaan antar anggota dalam satu cluster.
- **Heterogenitas external** (between cluster); yaitu perbedaan antara cluster yang satu dengan cluster yang lain.

Langkah pengelompokan dalam analisis cluster mencakup 3 hal berikut :

1. Mengukur kesamaan jarak
2. Membentuk cluster secara hirarkis
3. Menentukan jumlah cluster.

Adapun **metode pengelompokan** dalam analisis cluster meliputi :

- **Metode Hirarkis**; *memulai pengelompokan dengan dua atau lebih obyek yang mempunyai kesamaan paling dekat*. Kemudian diteruskan pada obyek yang lain dan seterusnya hingga cluster akan membentuk semacam 'pohon' dimana terdapat tingkatan (hirarki) yang jelas antar obyek, dari yang paling mirip hingga yang paling tidak mirip. Alat yang membantu untuk memperjelas proses hirarki ini disebut "**dendogram**".
- **Metode Non-Hirarkis**; *dimulai dengan menentukan terlebih dahulu jumlah cluster yang diinginkan* (dua, tiga, atau yang lain). Setelah jumlah cluster ditentukan, maka proses cluster dilakukan dengan

tanpa mengikuti proses hirarki. Metode ini biasa disebut "**K-Means Cluster**".

Asumsi yang harus dipenuhi dalam Analisis Cluster yaitu :

- Sampel yang diambil benar-benar dapat mewakili populasi yang ada (**representativeness of the sample**)
- **Multikolinieritas**.

Contoh Kasus

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

- jumlah pendapatan kota (trilyun Rp)
- jumlah pinjaman pemerintah kota (milyar Rp)
- jumlah dana hibah yang dimiliki kota (milyar Rp)
- jumlah konsumsi pemerintah kota (milyar Rp)
- jumlah penduduk kota (juta jiwa).

Untuk itu data yang berhasil dikumpulkan sebagai berikut:

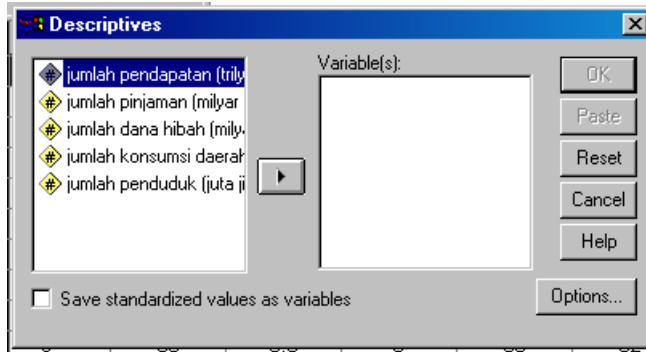
	kota	pendptan	pinjaman	dn.hibah	konsumsi	pduduk
1	A	55	5.6	9	53	25
2	B	61	8.0	7	62	41
3	C	58	3.9	7	61	32
4	D	67	5.5	7	64	51
5	E	71	5.7	6	73	42
6	F	76	7.6	8	83	29
7	G	81	8.7	9	83	57
8	H	56	7.1	6	85	29
9	I	84	7.6	7	82	46
10	J	88	6.5	8	85	52
11	K	84	6.8	9	83	61
12	L	90	8.0	9	93	66

Untuk menyelesaikan contoh kasus di atas dengan menggunakan aplikasi program SPSS, perlu dilakukan beberapa langkah berikut ini.

Langkah - 1 : "Standardisasi/Transformasi"

Mengingat data yang terkumpul mempunyai variabilitas satuan, maka perlu dilakukan langkah standardisasi atau transformasi terhadap variabel yang relevan ke bentuk z-score, sebagai berikut :

- 1) Setelah keseluruhan data yang dikumpulkan tersebut diatas dientry dalam program SPSS, selanjutnya klik menu "**analyze**" dan pilih sub menu "**Descriptives Statistics**" lalu "**Descriptives**" hingga muncul tampilan berikut ini :



2) Masukkan ke dalam kotak **VARIABLES** seluruh variabel instrumen penilai, yaitu variabel jumlah pendapatan, jumlah pinjaman, jumlah dana hibah, jumlah konsumsi, dan jumlah penduduk. (dalam hal ini variabel kota tidak dimasukkan karena data bertipe string). Kemudian aktifkan bagian "**Save standardized values as variables**". Abaikan bagian yang lain lalu tekan **OK** untuk menampilkan output aplikasi program SPSS.

Output yang didapat yaitu deskripsi dari keseluruhan variabel yang meliputi nilai maksimum, nilai minimum, rata-rata, dan standar deviasi dari masing-masing variabel.

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
jumlah pendapatan (trilyun Rp)	12	55	90	72.58	12.965
jumlah pinjaman (milyar Rp)	12	3.9	8.7	6.750	1.3708
jumlah dana hibah (milyar Rp)	12	6	9	7.67	1.155
jumlah konsumsi daerah (milyar Rp)	12	50	90	74.83	13.141
jumlah penduduk (juta jiwa)	12	25	66	44.25	13.572
Valid N (listwise)	12				

Namun, deskripsi tersebut diatas digunakan sebagai dasar perhitungan z-score yang diperoleh. Selanjutnya buka tampilan "**data view**" dari tabel data. Hal yang akan dijumpai sebagai berikut :

	zpendpta	zpinjama	zdn.hiba	zkonsums	zpduduk
1	-1.35625	-.83893	1.15470	-1.88970	-1.41834
2	-.89345	.91188	-.57735	-.97656	-.23946
3	-1.12485	-2.07908	-.57735	-1.12875	-.90258
4	-.43066	-.91188	-.57735	-.82437	.49734
5	-.12213	-.76598	-1.44338	-.36779	-.16578
6	.26354	.62008	.28868	.39316	-1.12362
7	.64920	1.42253	1.15470	.39316	.93942
8	-1.27911	.25533	-1.44338	.84973	-1.12362
9	.88060	.62008	-.57735	.54535	.12894
10	1.18913	-.18238	.28868	.84973	.57102
11	.88060	.03648	1.15470	1.00192	1.23414
12	1.34339	.91188	1.15470	1.15411	1.60254

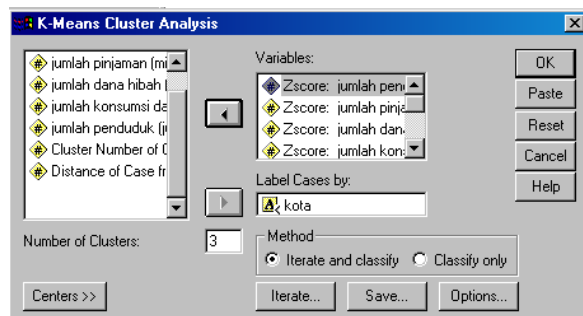
Untuk selanjutnya, hasil z-score inilah yang akan dipakai dasar analisis cluster. Namun apabila data yang terkumpul tidak mempunyai variabilitas satuan, maka prose analisis cluster dapat langsung dilakukan tanpa terlebih dahulu melakukan transformasi atau standardisasi.

Langkah - 2 : "Analisis Cluster"

A. Metode K-Means Cluster (Non-Hirarkis)

Sebagaimana telah dijelaskan sebelumnya bahwa metode K-Means Cluster ini jumlah cluster ditentukan sendiri. Oleh karena itu, berikut ini langkah-langkah yang harus dilakukan dalam menggunakan metode K-Means Cluster dalam aplikasi program SPSS. Perlu diingat bahwa bahan analisis bukan lagi data asli, namun data hasil transformasi/standardisasi.

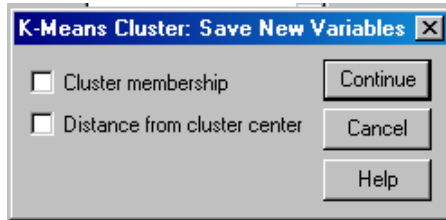
- 1) Dari tampilan data yang tertera (hasil standardisasi/transformasi), buka menu "**Analyze**", lalu pilih sub menu "**Classify**" dan pilih "**K-Means Cluster...**" hingga tampak pada layar sebagai berikut :



- 2) Masukkan **seluruh variabel Z-Score** ke dalam kotak **VARIABLES**. Kemudian variabel **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

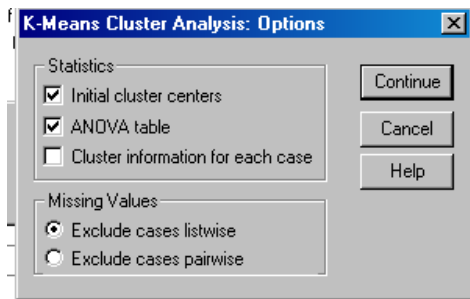
ini diisi 3, berarti diharapkan akan dibentuknya 3 cluster.

- 3) Kemudian klik mouse pada kotak "Save..." hingga muncul tampilan seperti berikut ini :



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

- 4) Aktifkan kedua kotak dalam menu Save, yaitu "**Cluster membership**" dan "**Distance from cluster center**". Selanjutnya tekan tombol "**Continue**" untuk kembali ke menu utama.
- 5) Kemudian klik mouse pada kotak "**Options...**" hingga tampak tampilan berikut ini :



Pada bagian **Statistics**, aktifkan "**Initial cluster centers**" dan "**ANOVA table**". Abaikan bagian yang lain, lalu tekan "**Continue**" untuk kembali ke menu utama.

- 6) Dari tampilan menu utama cluster, abaikan bagian yang lain lalu tekan tombol **OK** untuk dapat menampilkan output aplikasi program SPSS seperti berikut ini.

Initial Cluster Centers

	Cluster		
	1	2	3
Zscore: jumlah pendapatan (trilyun Rp)	1.34339	-1.27911	-1.12485
Zscore: jumlah pinjaman (milyar Rp)	.91188	.25533	-2.07908
Zscore: jumlah dana hibah (milyar Rp)	1.15470	-1.44338	-.57735
Zscore: jumlah konsumsi daerah (milyar)	1.15411	.84973	-1.12875
Zscore: jumlah penduduk (juta jiwa)	1.60254	-1.12362	-.90258

Tabel diatas 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^a

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 distance change. The maximum distance by which any center has changed is .000. The current iteration is 2. The minimum distance between initial centers is 3.192.

Ternyata proses clustering yang dilakukan melalui 2 tahapan iterasi untuk mendapatkan cluster yang tepat. Dari tabel diatas 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: jumlah pendapatan (trilyun Rp)	.98858	-.63634	-.75847
Zscore: jumlah pinjaman (milyar Rp)	.56172	.59576	-1.14896
Zscore: jumlah dana hibah (milyar Rp)	.63509	-.57735	-.36084
Zscore: jumlah konsumsi daerah (milyar)	.78885	.08878	-1.05265
Zscore: jumlah penduduk (juta jiwa)	.89521	-.82890	-.49734

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

- Nilai negatif (-) berarti data berada di bawah rata-rata total.

- Nilai positif (+) berarti data berada di atas rata-rata total.

Rumus umum yang digunakan yaitu :

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

Dimana :

- X : rata-rata sampel (variabel dalam cluster)
- μ : rata-rata populasi
- Z : nilai standardisasi
- σ : standar deviasi

Sebagai contoh, apabila ingin diketahui rata-rata jumlah pendapatan kota di cluster-1 yaitu :

$$\begin{aligned} & (\text{rata-rata pendapatan seluruh kota}) + (0,98858 \times \text{standar deviasi rata-rata pendapatan}) \\ & = 72,58 + (0,98858 \times 12,965) \\ & = 85,3969 \end{aligned}$$

Jadi rata-rata jumlah pendapatan kota yang berada di cluster-1 adalah Rp 85,3969 trilyun.

Demikian seterusnya dapat diketahui rata-rata nilai masing-masing variabel dalam tiap cluster.

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

- **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.

- **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 variabel 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.

▪ **Cluster-3**

Sedangkan 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.

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

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

ANOVA

	Cluster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
Zscore: jumlah pendapatan (trilyun Rp)	4.201	2	.289	9	14.556	.002
Zscore: jumlah pinjaman (milyar Rp)	3.961	2	.342	9	11.587	.003
Zscore: jumlah dana hibah (milyar Rp)	1.769	2	.829	9	2.133	.174
Zscore: jumlah konsumsi daerah (milyar)	3.784	2	.381	9	9.921	.005
Zscore: jumlah penduduk (juta jiwa)	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 :

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

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 (sig < 0,05), maka semakin besar perbedaan variabel pada cluster yang terbentuk.

Dengan demikian hasil cluster yang didapat dalam penelitian ini bahwa untuk instrumen jumlah pendapatan kota yang paling menunjukkan adanya perbedaan diantara kota-kota pada ketiga cluster yang terbentuk. Hal ini dengan ditunjukkannya nilai $F = 14,556$ dan $sig = 0,002$. Dan untuk variabel 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

Nampak jelas bahwa cluster-1 beranggotakan 5 kota, cluster-2 berisi 3 kota, dan pada cluster-3 terdapat 4 kota yang mengelompok. Dan untuk mengetahui kota-kota mana saja yang masuk dalam kategori tiap-tiap cluster dapat kembali dibuka tampilan "**data view**" pada kolom terakhir akan nampak seperti berikut ini :

	kota	pend	pinja	dn.hj	kon	pdu	zpen	zpinj	zdn	zko	zpd	qcl_1	qcl_2
1	A	55	5.6	9	50	25	-1.4	-.8	1.2	-2	-1	3	2.07346
2	B	61	8.0	7	62	41	-.89	.91	-.6	-1	-.2	2	1.28390
3	C	58	3.9	7	60	32	-1.1	-2	-.6	-1	-.9	3	1.10283
4	D	67	5.5	7	64	51	-.43	-.9	-.6	-.8	.50	3	1.11895
5	E	71	5.7	6	70	42	-.12	-.8	-1	-.4	-.2	3	1.51738
6	F	76	7.6	8	80	29	.264	.62	.29	.39	-1	2	1.31905
7	G	81	8.7	9	80	57	.649	1.4	1.2	.39	.94	1	1.13345
8	H	56	7.1	6	86	29	-1.3	.26	-1	.85	-1	2	1.39462
9	I	84	7.6	7	82	46	.881	.62	-.6	.55	.13	1	1.45998
10	J	88	6.5	8	86	52	1.19	-.2	.29	.85	.57	1	.90703
11	K	84	6.8	9	88	61	.881	.04	1.2	1.0	1.2	1	.84724
12	L	90	8.0	9	90	66	1.34	.91	1.2	1.2	1.6	1	1.07342

Perhatikan 2 kolom terakhir pada tabel di atas. "**qcl_1**" menunjukkan **nomor cluster** dari keberadaan kota, dan "**qcl_2**" merupakan **jarak antara obyek dengan pusat cluster**. Dengan demikian, dapat ditafsirkan sebagai berikut :

- **Cluster-1** : berisikan kota **G, I, J, K, dan L** dengan masing-masing jarak terhadap pusat cluster-1 adalah 1,13345; 1,45998; 0,90703; 0,84724; dan 1,07342.

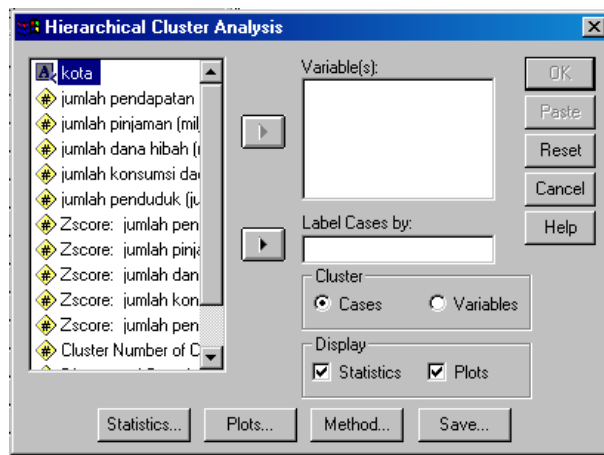
- **Cluster-2** : berisikan kota **B, F, dan H**, dengan masing-masing jarak terhadap pusat cluster-2 adalah 1,28390; 1,31905; dan 1,39462.
- **Cluster-3** : berisikan kota **A, C, D, dan E**, dengan masing-masing jarak terhadap pusat cluster-3 adalah 2,07346; 1,10283; 1,11895; dan 1,51738.

B. Metode Hierarchical Cluster (Hirarkis)

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 aglomerativ**" yang digambarkan dengan dendogram.

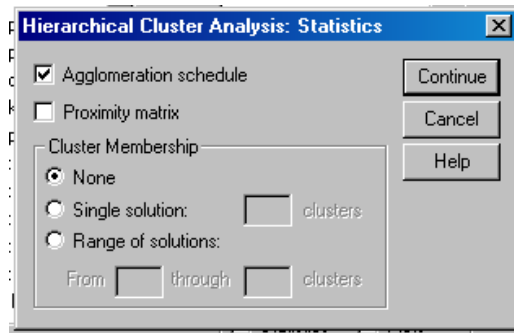
Contoh kasus di atas akan dicoba untuk diselesaikan pula dengan metode aglomerativ. Untuk itu, langkah-langkah yang harus dilakukan dalam aplikasi program SPSS sebagai berikut :

- 1) Masih dengan data sebelumnya, yang merupakan hasil standardisasi, buka menu "**Analyze**" lalu pilih sub menu "**Classify**" kemudian "**Hierarchical Cluster...**" hingga muncul tampilan seperti berikut ini :



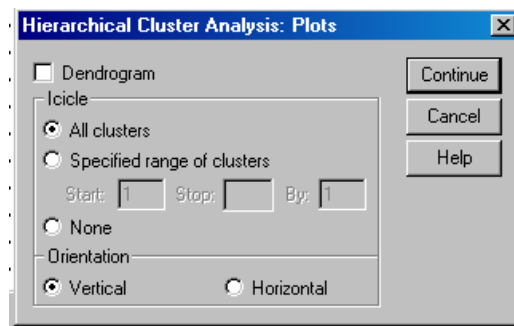
- 2) Masukkan seluruh variabel yang telah distandardkan (Z-score) ke dalam bagian "**Variable(s)**". Untuk bagian "**Label Cases by**" isi dengan variabel **kota**; sedangkan untuk bagian "**Cluster**" pilih **Cases**; pada bagian "**Display**" pilih keduanya yaitu **Statistics** dan **Plots**.

- 3) Kemudian klik mouse pada kotak **"Statistics"** hingga muncul tampilan berikut ini :



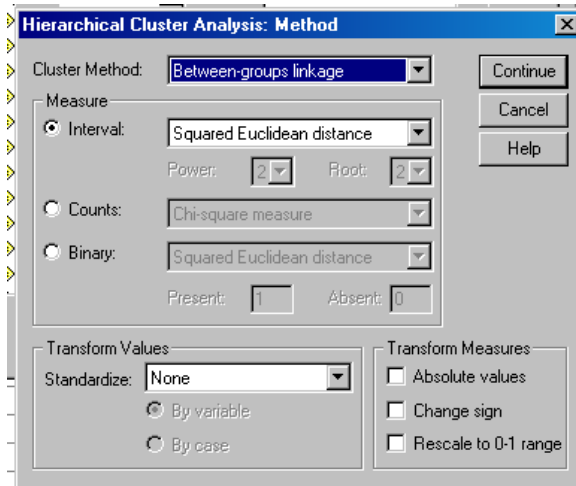
Selain kotak **"Agglomeration Schedule"**, aktifkan pula kotak **"Proximity matrix"** untuk menampilkan jarak antar variabel. Pada bagian **"Cluster Membership"** klik mouse pada pilihan **Range of Solutions** lalu ketik **2** pada **FROM** dan **4** pada **THROUGH** (berarti nantinya akan ditampilkan susunan 2, 3, dan 4 cluster). Kemudian tekan tombol **"Continue"** untuk kembali ke menu utama.

- 4) Selanjutnya klik mouse pada kotak **"Plots"** hingga muncul tampilan seperti berikut ini :



Aktifkan pilihan **"Dendrogram"**; kemudian pada bagian **"Icicle"** pilih **None**. Abaikan bagian yang lain lalu tekan tombol **"Continue"** untuk kembali ke menu utama.

- 5) Berikutnya klik mouse pada kotak **"Method..."** hingga muncul tampilan :



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** untuk menampilkan output aplikasi program SPSS seperti berikut.

Case Processing Summary^a

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 diatas menunjukkan matrik jarak antara variabel satu dengan variabel yang lain. **Semakin kecil jarak euclidean,**

maka semakin mirip kedua variabel 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 anatar varaiebl 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
 - Jarak rata-rata = $(2,432 + 1,761) / 2 = 2,0965$
- Stage 2 : terjadi pembentukan cluster Kota D dan Kota E (berjarak 1,515), yang kemudian berlanjut ke stage 6.
- 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 diatas dapat dijabarkan bahwa :

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

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

- Apabila ditentukan dibentuk 2 cluster, maka :
 - Anggota cluster 1 adalah Kota A, B, C, D, E, dan H
 - Anggota cluster 2 adalah Kota F, G, I, J, K, dan L.

selanjutnya lakukan analisis cluster dengan menggunakan metode :

a) **K-Means Cluster** (untuk mahasiswa **NRP Ganjil**)

- Sebutkan anggota masing-masing cluster dari 4 cluster yang terbentuk.
- Hitung rata-rata harga dari persepsi responden terhadap Supermarket, di keempat cluster yang terbentuk.

b) **Hierarchical Cluster** (untuk mahasiswa **NRP Genap**)

- Sebutkan anggota masing-masing cluster dari 4 cluster yang terbentuk.
- Sebutkan obyek/responden yang memiliki jarak euclidean terkecil.

Catatan :

- Perlu diingat bahwa data masih belum ditransformasi.