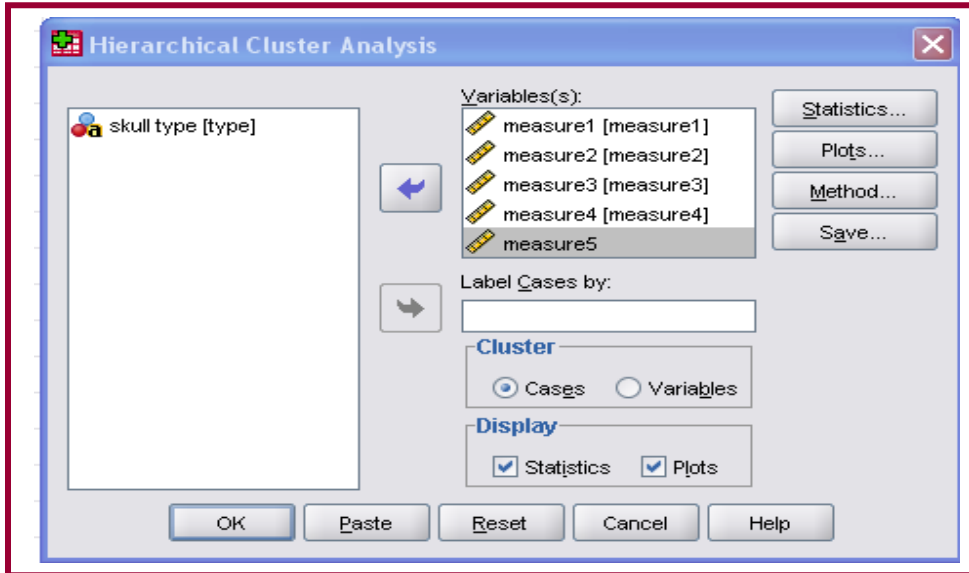


## Cluster Analysis Workshop#1-Hierarchical Clustering Method

ภาพด้านล่างแสดงข้อมูลที่ใช้ในCA โดย measure1 ถึง measure 5 เป็นการวัดขนาดของกะโหลกศีรษะในมุมมองห้าด้าน

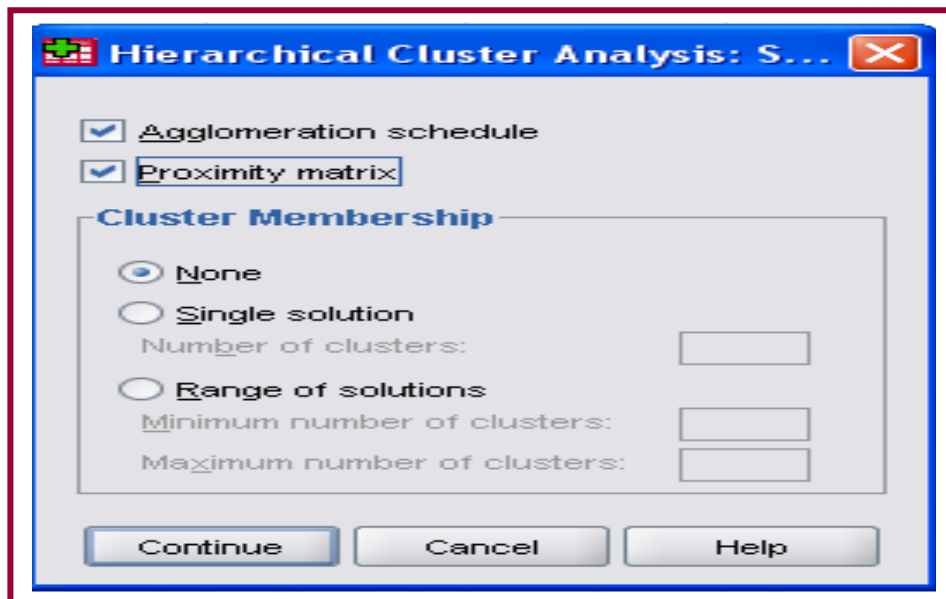
	type	measure1	measure2	measure3	measure4	measure5
1	A	190.5	152.5	145.0	73.5	136.5
2	A	172.5	132.0	125.5	63.0	121.0
3	A	167.0	130.0	125.5	69.5	119.5
4	A	169.5	150.5	133.5	64.5	128.0
5	A	175.0	138.5	126.0	77.5	135.5
6	A	177.5	142.5	142.5	71.5	131.0
7	A	179.5	142.5	127.5	70.5	134.5
8	A	179.5	138.0	133.5	73.5	132.5
9	A	173.5	135.5	130.5	70.0	133.5
10	A	162.5	139.0	131.0	62.0	126.0
11	A	178.5	135.0	136.0	71.0	124.0
12	A	171.5	148.5	132.5	65.0	146.5
13	A	180.5	139.0	132.0	74.5	134.5
14	A	183.0	149.0	121.5	76.5	142.0
15	A	169.5	130.0	131.0	68.0	119.0
16	A	172.0	140.0	136.0	70.5	133.5
17	A	170.0	126.5	134.5	66.0	118.5
18	B	182.5	136.0	138.5	76.0	134.0
19	B	179.5	135.0	128.5	74.0	132.0
20	B	191.0	140.5	140.5	72.5	131.5
21	B	184.5	141.5	134.5	76.5	141.5
22	B	181.0	142.0	132.5	79.0	136.5
23	B	173.5	136.5	126.0	71.5	136.5
24	B	188.5	130.0	143.0	79.5	136.0
25	B	175.0	153.0	130.0	76.5	142.0
26	B	196.0	142.5	123.5	76.0	134.0
27	B	200.0	139.5	143.5	82.5	146.0
28	B	185.0	134.5	140.0	81.5	137.0
29	B	174.5	143.5	132.5	74.0	136.5
30	B	195.5	144.0	138.5	78.5	144.0
31	B	197.0	131.5	135.0	80.5	139.0
32	B	182.5	131.0	135.0	68.5	136.0

- คลิก Analyze>Classify>Hierarchical cluster



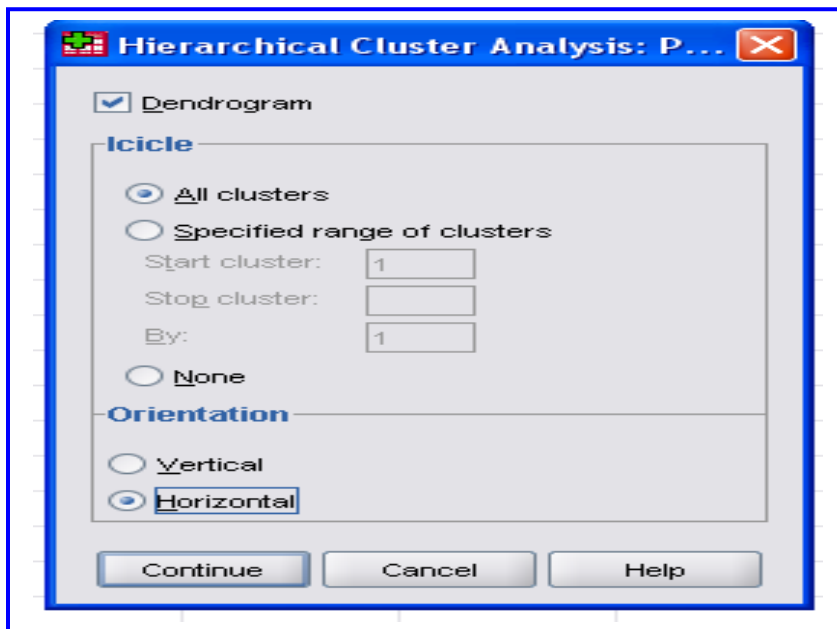
คลิกเลือกตัวแปร  
(measure1,...measure5)  
ให้ไปอยู่ในกล่อง variables

- คลิกที่ปุ่ม Statistics จะแสดงmenu ดังภาพด้านล่าง



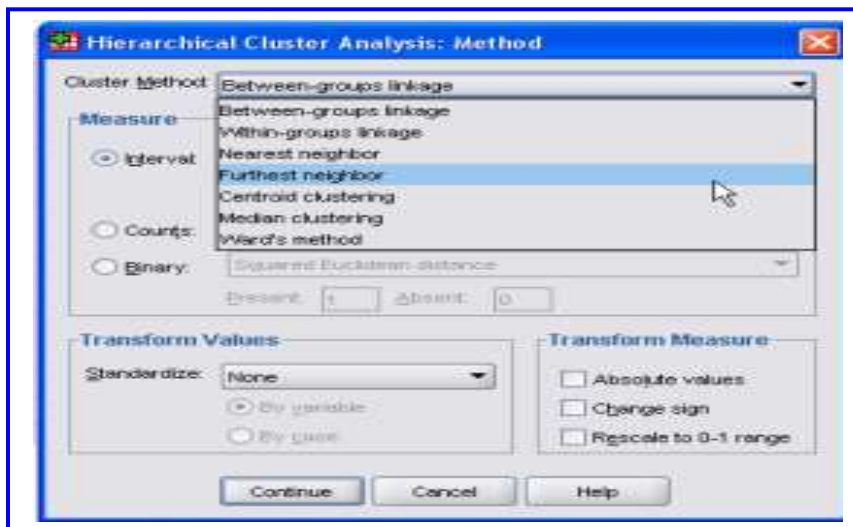
คลิกเลือก Agglomeration  
schedule และ Proximity  
matrix

- คลิกที่ปุ่ม Plots จะปรากฏ menu ด้านล่าง



คลิกเลือก Dendrogram ในส่วน  
Icicle คลิกเลือก Horizontal  
Orientation

- คลิกที่ปุ่ม Method จะปรากฏ menu ด้านล่าง

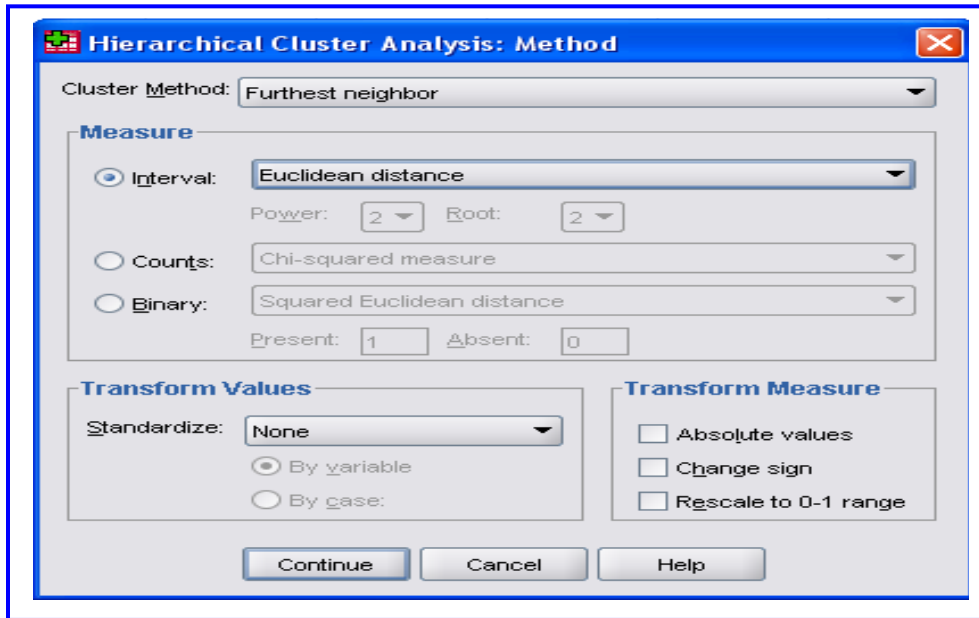


คลิกเลือกการรวมกลุ่มแบบ  
furthest neighbor

ในส่วนของมาตรวัดระยะทางเพื่อเปรียบเทียบความเหมือน/ แยกแยะความแตกต่าง(measure) เลือกใช้

Euclidean distance เนื่องจากทั้งห้าตัวแปรใช้หน่วยวัด( มิลลิเมตร )เหมือนกันและมี standard deviation

ไม่แตกต่างกัน



- คลิก Continue ภาพด้านล่างแสดง computer output เริ่มต้นด้วย Proximity Matrixบางส่วน(ที่แสดงระยะห่างระหว่างcase แต่ละ case)

Case	Proximity Matrix																							
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	.000	38.496	41.759	37.020	28.556	17.593	23.343	21.997	28.452	37.480	26.329	28.581	21.368	25.890	38.468	34.444	39.834	19.774	36.633	13.802	17.357	19.672	20.166	23.483
2	38.496	.000	9.874	21.599	21.662	24.423	20.044	18.796	15.804	14.327	15.108	31.249	21.674	32.226	8.500	18.872	11.522	24.980	18.578	28.963	30.294	26.796	18.276	32.811
3	41.759	8.874	.000	24.244	21.366	28.381	23.282	21.862	17.136	15.186	17.029	34.670	23.568	34.471	8.245	28.791	10.747	26.415	19.151	32.623	32.442	27.717	18.436	33.775
4	27.020	21.599	24.244	.000	21.348	18.340	16.681	18.900	17.621	14.862	16.641	18.740	19.944	25.855	22.798	13.748	25.880	23.627	21.714	26.229	25.164	22.891	19.736	34.066
5	28.525	21.662	21.366	21.348	.000	18.728	6.407	18.887	9.578	32.827	17.263	28.779	8.748	15.419	22.136	12.816	25.738	14.842	7.963	22.811	14.440	8.882	6.576	23.464
6	17.593	24.423	26.301	18.340	18.728	.000	15.584	16.559	14.748	22.817	12.186	21.319	12.358	25.881	22.556	8.274	23.722	10.687	16.233	13.838	15.788	14.937	18.828	19.144
7	23.343	20.044	23.282	16.681	6.407	15.584	.000	8.222	8.760	31.491	15.482	17.299	7.026	13.919	22.894	11.851	25.913	14.239	9.703	17.943	12.848	18.186	8.902	23.680
8	21.997	18.796	21.662	18.908	10.087	18.559	6.322	.000	8.031	21.897	9.734	21.059	3.041	18.404	19.564	8.746	21.812	8.819	5.074	13.766	11.325	8.893	10.700	16.637
9	28.452	15.804	17.136	17.621	9.578	14.748	9.780	8.031	.000	15.930	12.114	19.381	9.206	21.833	18.148	7.380	18.715	13.472	7.648	21.012	16.711	13.874	5.701	22.538
10	37.480	14.327	15.186	14.862	22.627	22.817	21.401	21.897	15.930	.000	19.545	24.551	23.527	32.914	14.863	15.644	17.234	26.930	22.164	32.300	30.871	27.436	18.782	36.156
11	26.329	15.108	17.029	19.944	17.262	12.186	15.482	8.734	12.114	19.545	.000	28.031	12.586	27.842	12.845	12.580	14.213	12.178	11.413	16.288	20.421	16.963	18.845	19.755
12	28.581	31.249	34.070	18.748	20.779	21.319	17.288	21.859	19.261	24.551	28.031	.000	20.143	28.149	33.373	16.852	35.711	24.380	23.527	28.089	19.474	26.700	18.236	32.588
13	21.368	21.674	23.568	19.944	8.746	12.359	7.628	3.041	8.206	32.527	12.580	20.143	.000	18.836	22.034	18.308	24.515	7.599	5.978	14.862	9.828	5.809	10.210	17.125
14	25.890	32.226	34.471	25.855	15.419	25.881	12.518	18.404	21.833	32.814	27.942	30.149	18.638	.000	35.136	22.814	38.817	22.858	19.066	24.970	15.091	14.509	17.944	29.975
15	38.489	8.500	6.245	22.798	22.136	22.555	22.684	19.564	16.148	14.883	12.845	33.373	22.034	25.139	.000	18.648	5.385	23.457	18.337	28.570	38.796	28.566	20.044	36.434
16	24.444	18.872	20.791	13.748	12.816	8.274	11.851	8.746	7.280	15.844	12.580	16.852	10.308	22.914	18.848	.000	20.827	12.787	12.328	19.738	18.148	13.380	11.158	22.550
17	39.834	11.522	10.747	25.880	25.738	23.722	25.913	21.812	18.715	17.234	14.213	35.711	24.515	38.817	5.385	28.827	.000	24.551	21.088	28.735	22.780	28.296	23.211	30.253
18	19.774	24.860	26.415	23.927	14.942	18.697	14.238	6.819	13.472	38.930	12.178	24.300	7.588	22.858	23.457	12.757	24.951	.000	10.863	10.724	10.332	8.480	16.248	10.416
19	26.833	17.578	19.151	21.714	7.963	18.233	8.783	5.874	7.648	22.184	11.413	32.527	5.978	18.885	18.337	22.329	21.888	10.863	.000	17.578	14.133	18.837	8.428	19.038
20	13.802	28.963	32.623	26.220	22.611	13.636	17.848	13.786	21.812	32.330	16.288	28.088	14.062	24.970	28.970	18.736	29.736	10.724	17.578	.000	13.973	15.291	23.833	13.856
21	17.357	30.294	32.442	25.164	14.448	15.788	12.648	11.325	16.711	30.871	20.421	19.474	9.028	15.891	30.790	16.140	32.700	10.332	14.133	13.973	.000	6.910	16.378	16.117
22	18.672	26.796	27.717	22.891	9.682	14.187	10.188	8.893	13.874	27.426	16.963	20.700	5.888	14.509	28.568	13.380	29.286	9.480	10.607	15.281	8.910	.000	13.601	17.835
23	30.186	18.276	19.438	18.736	6.576	18.828	8.902	10.700	5.701	18.792	16.845	16.235	10.216	17.944	20.044	11.158	22.211	16.248	8.428	23.833	18.378	13.601	.000	24.910
24	23.483	32.811	33.775	34.066	23.464	18.144	23.888	16.837	22.538	36.156	19.755	32.588	17.125	28.875	30.434	22.850	30.253	10.416	19.038	13.858	15.117	17.835	24.910	.000
25	22.467	33.827	34.183	18.717	16.416	28.483	15.123	18.889	20.573	38.827	26.964	13.829	17.044	12.339	34.061	17.951	37.550	21.949	21.278	25.471	15.588	14.133	18.594	30.418
26	24.587	31.894	35.408	32.184	21.628	27.885	17.854	28.825	25.308	38.989	25.372	31.536	18.088	16.718	34.895	27.721	37.326	21.281	19.013	18.244	17.628	17.830	23.976	24.678
27	20.883	48.279	48.728	42.418	32.613	28.384	31.088	28.831	34.637	48.916	33.985	38.342	26.627	36.253	48.000	33.775	48.874	23.816	30.557	26.031	19.532	24.395	35.954	18.214
28	21.059	31.181	31.688	30.141	18.172	18.282	19.426	13.900	19.183	33.288	18.378	28.231	12.628	24.827	29.116	18.888	30.026	7.071	15.628	12.548	11.189	11.819	20.787	6.819

ตัวเลขที่ปรากฏใน Proximity Matrix จะแสดงระยะห่าง( Euclidean distance) ระหว่าง case คู่หนึ่งๆ ตัวอย่างเช่น

ระหว่างcase ที่ 1 และ 2 มีระยะห่างเท่ากับ  $38.406\{\text{square root ของ } (190.5-172.5)^2 + (152.5-132.0)^2 + (145-125.5)^2 + (73.5-63.0)^2 + (136.5-121)^2\}$

ส่วนต่อไปจะแสดง Agglomeration Scheduleในรูปแบบตาราง

**Complete**

**Agglomeration Schedule**

Stage #	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	8	13	3.041	0	0	4
2	15	17	5.385	0	0	14
3	9	23	5.701	0	0	11
4	8	19	5.979	1	0	8
5	24	28	6.819	0	0	17
6	21	22	6.910	0	0	21
7	16	29	7.211	0	0	15
8	7	8	8.703	0	4	13
9	2	3	8.874	0	0	14
10	27	30	9.247	0	0	23
11	5	9	9.579	0	3	13
12	18	32	9.874	0	0	18
13	5	7	10.700	11	8	24
14	2	15	11.522	9	2	28
15	6	16	12.104	0	7	22
16	14	25	12.339	0	0	21
17	24	31	13.528	5	0	23
18	11	18	13.537	0	12	22
19	1	20	13.802	0	0	26
20	4	10	14.062	0	0	28
21	14	21	15.588	16	6	25
22	6	11	16.302	15	18	24
23	24	27	18.554	17	10	27
24	5	6	18.828	13	22	29
25	12	14	20.700	0	21	30
26	1	26	24.597	19	0	27
27	1	24	25.269	26	23	30
28	2	4	25.880	14	20	29
29	2	5	26.930	28	24	31
30	1	12	36.342	27	25	31
31	1	2	48.816	30	29	0

Agglomeration Schedule แสดงให้เห็นการรวมกลุ่มกันในแต่ละขั้นตอน โดยจะแสดงการรวมกันของ case 2 cases

ขึ้นเป็น cluster โดยคอลลัมน์ Coefficients แสดงระยะห่างระหว่าง cases ที่มารวมกัน ส่วนคอลลัมน์สุดท้ายแสดงลำดับ

ขั้นตอนที่จะมีการ update case ที่เป็นสมาชิกใหม่เข้ามาในกลุ่ม ตัวอย่างเช่น case ที่ 8 และ case ที่ 13 เมื่อรวมกัน

เข้าเป็น cluster (อาจเรียกชื่อว่า cluster 8+13) แล้ว ต่อมาก็จะมี case ที่ 19 เข้ามารวมด้วยในขั้นตอนที่ 4 กลายเป็น cluster

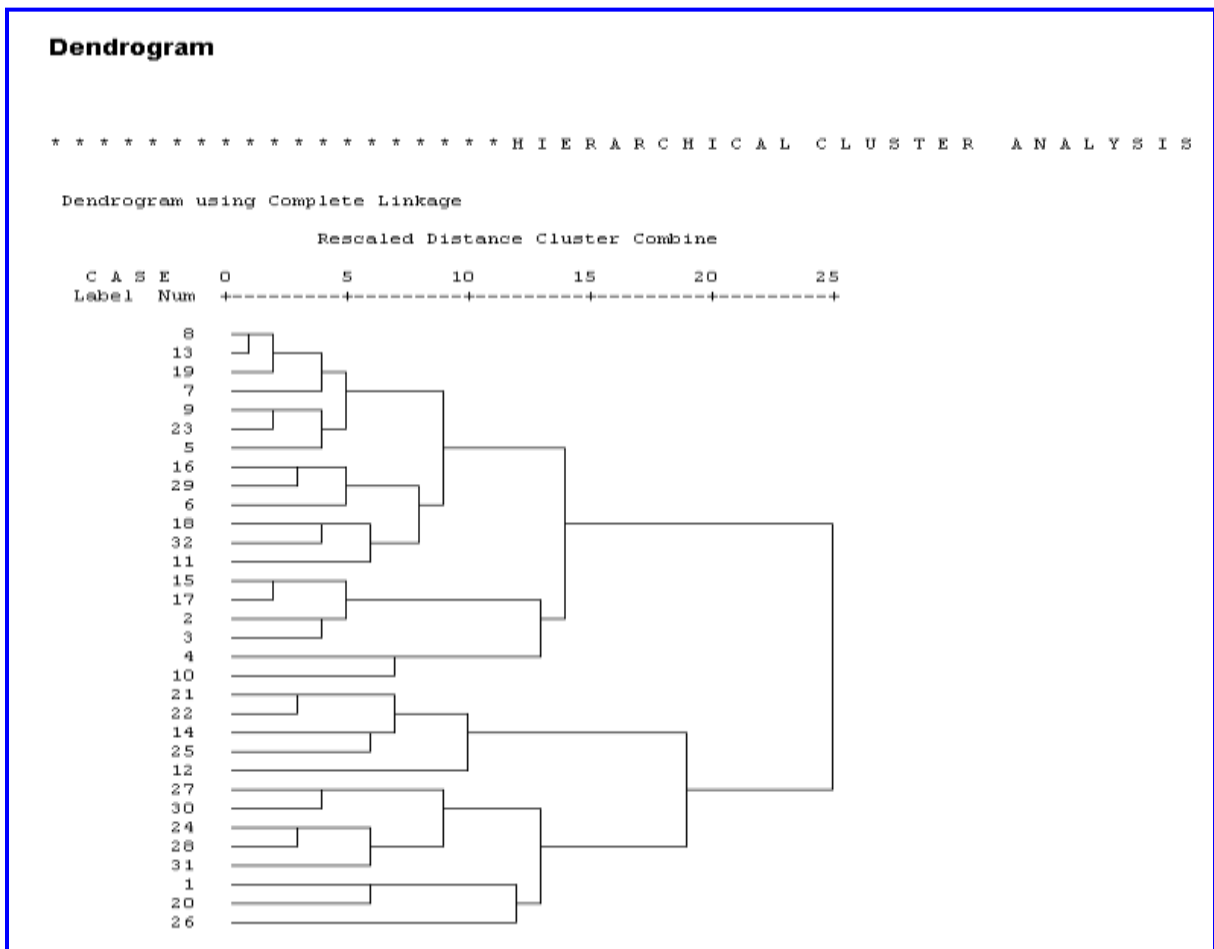
ที่ใหญ่ขึ้น ( อาจเรียกชื่อใหม่เป็น cluster 8+13+19)

ตารางต่อไปแสดง (horizontal ) icicle plot ( บางส่วน)

Horizontal icicle																												
Case	Number of clusters																											
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
32	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
18	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
11	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
29	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
16	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
6	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
19	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
13	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
8	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
7	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
23	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
9	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
5	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
10	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
4	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
17	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
15	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
3	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x

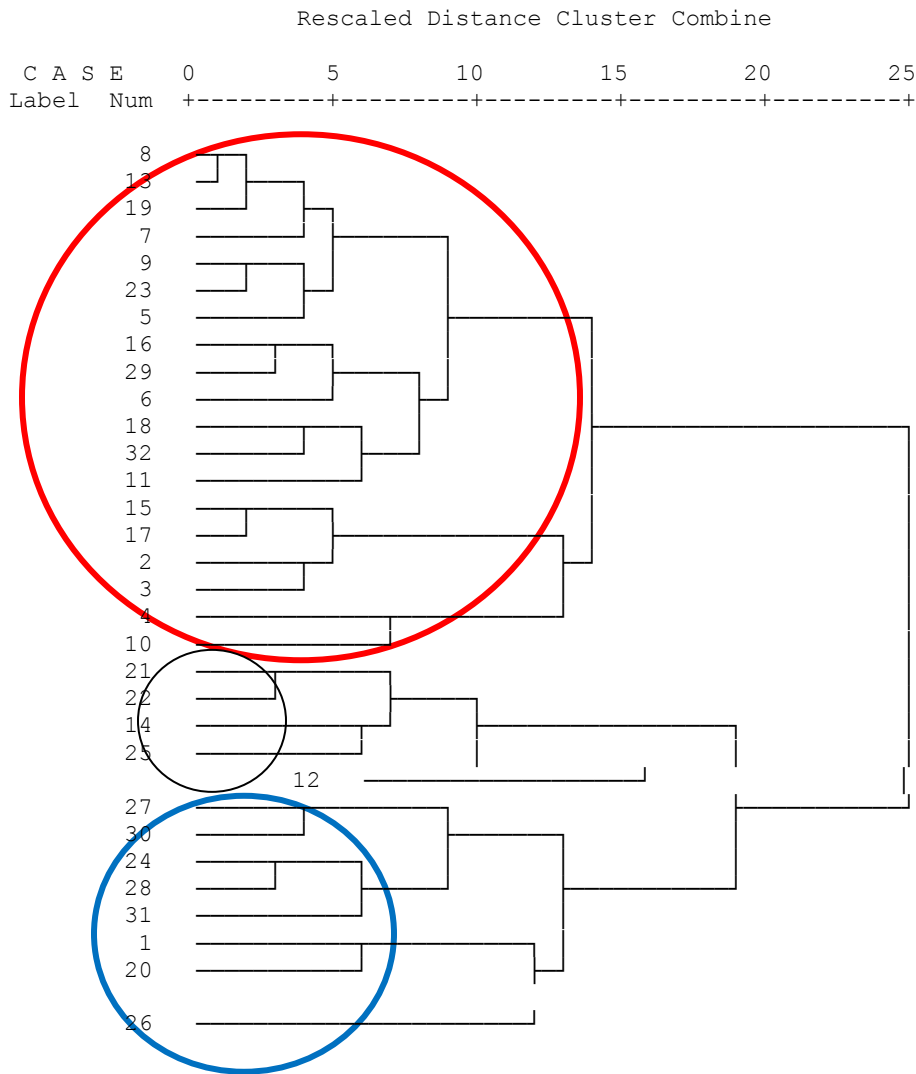
Icicle plot แสดงให้เห็นลำดับขั้นตอนการรวมกันเป็น cluster ในกรณีของ horizontal icicle plot เราต้องอ่านตารางเริ่มจากคอลัมน์ขวาสุดไปทางซ้าย ในตัวอย่างนี้ แสดงให้เห็นว่า case ที่ 13 และ 8 รวมกันเข้าเป็น cluster (อาจเรียกชื่อใหม่ว่า cluster13+8) ส่วน case อื่นๆแยกกันเป็น cluster ที่มีสมาชิกเพียงหนึ่ง case จากนั้น case ที่ 17 และ 15 รวมกันเข้าเป็น cluster (อาจเรียกชื่อใหม่ว่า cluster 17+15) ส่วน case อื่นๆยังแยกกันเป็น cluster ที่มีสมาชิกเพียงหนึ่ง case (ยกเว้น cluster 13+8) ลำดับต่อมา case ที่ 23 และ 9 รวมกันเข้าเป็น cluster (อาจเรียกชื่อใหม่ว่า cluster 23+9) เป็นเช่นนี้ไปเรื่อยๆ จนในที่สุดคอลัมน์ทางซ้ายสุดแสดงให้เห็น case ทุก case รวมกันเข้าเป็น cluster เดียว (เรียกชื่อว่า cluster 1+2+3+.....+32)

ส่วนต่อไปแสดงภาพของ Dendrogram



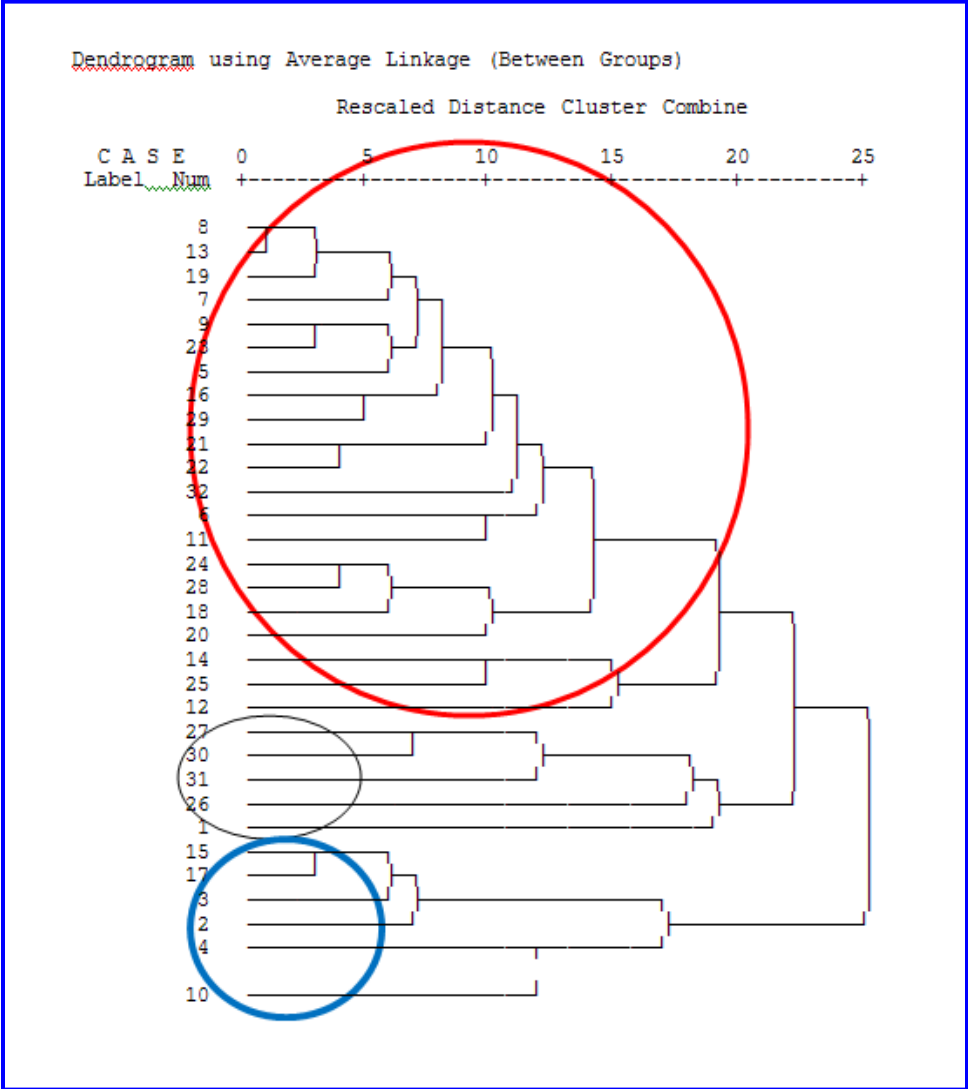
ซึ่งหากพิจารณาแล้ว case ต่าง ๆ สามารถจัดเข้าได้เป็น 2-3 cluster ภาพด้านล่างแสดงกรณีของ 3 cluster solution

Dendrogram using Complete Linkage



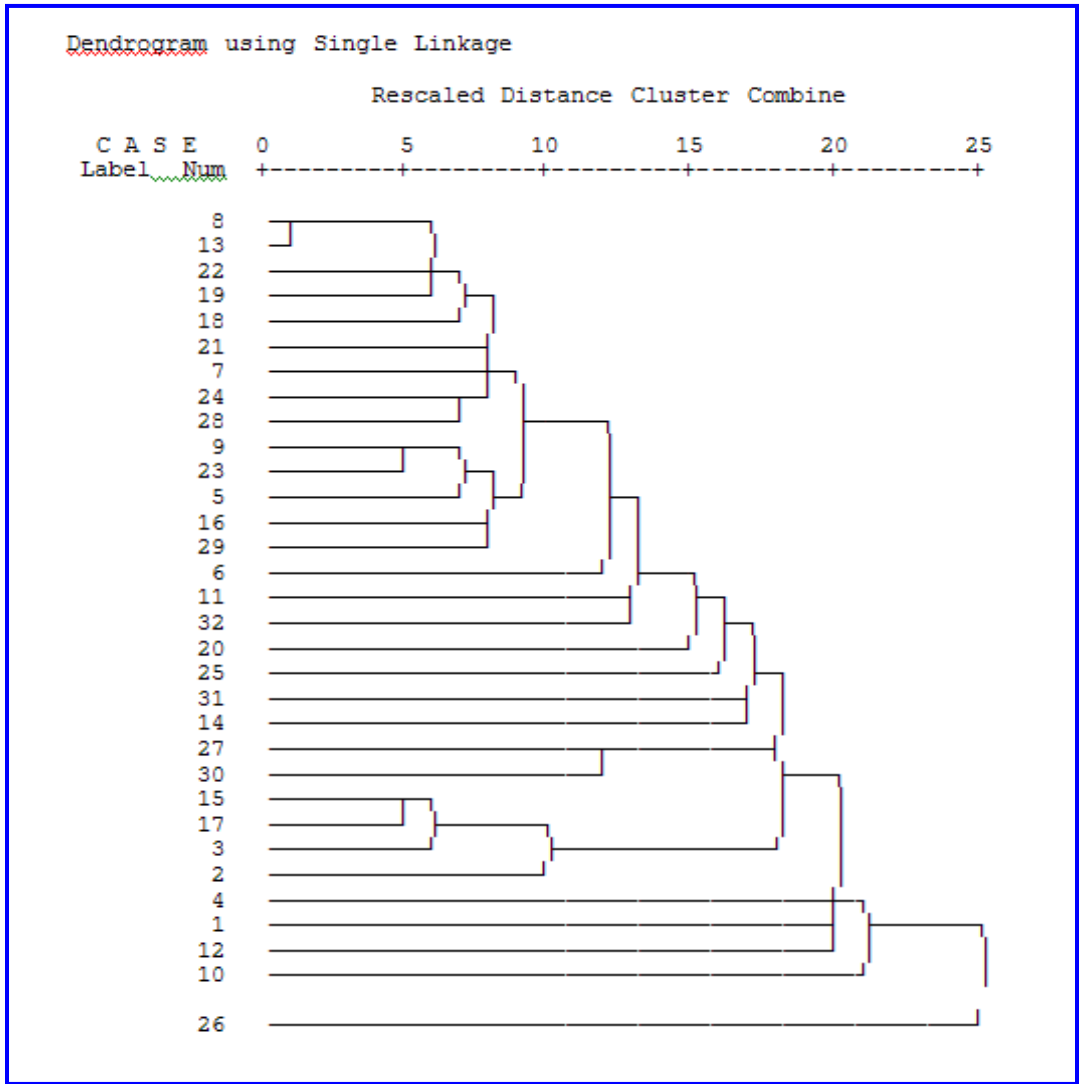
### Hierarchical clustering โดยใช้ Average Linkage Method

Dendrogram ที่ได้สามารถแสดงให้เห็นดังภาพด้านล่าง ซึ่งบอกเป็นนัยได้ว่า การจัด case ต่าง ๆ ทั้ง 32 cases เข้าเป็น cluster 3 cluster น่าจะมีความเหมาะสม



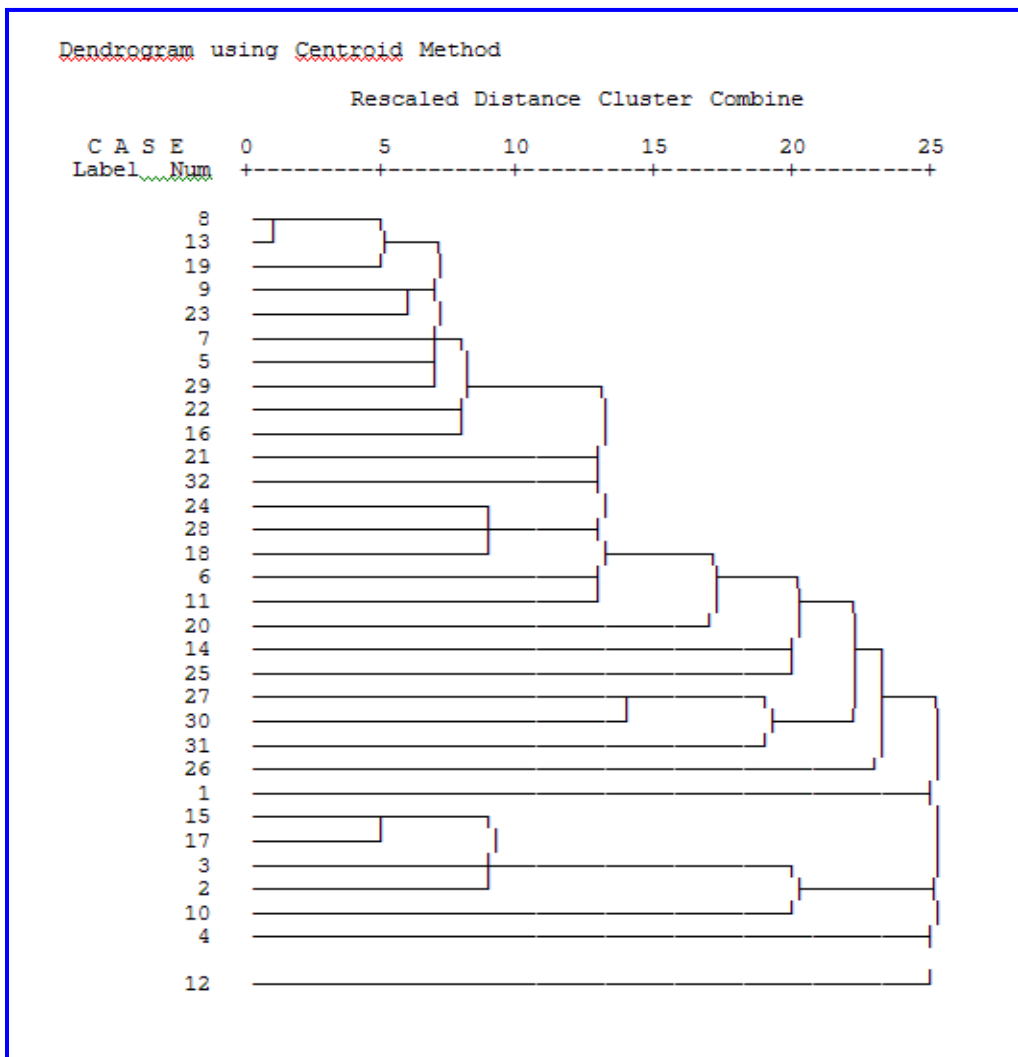
Hierarchical clustering โดยใช้ Nearest neighbor method

Dendrogram ในกรณีนี้มีแสดงให้เห็นในหน้าถัดไป



Hierarchical clustering โดยใช้ centroid method

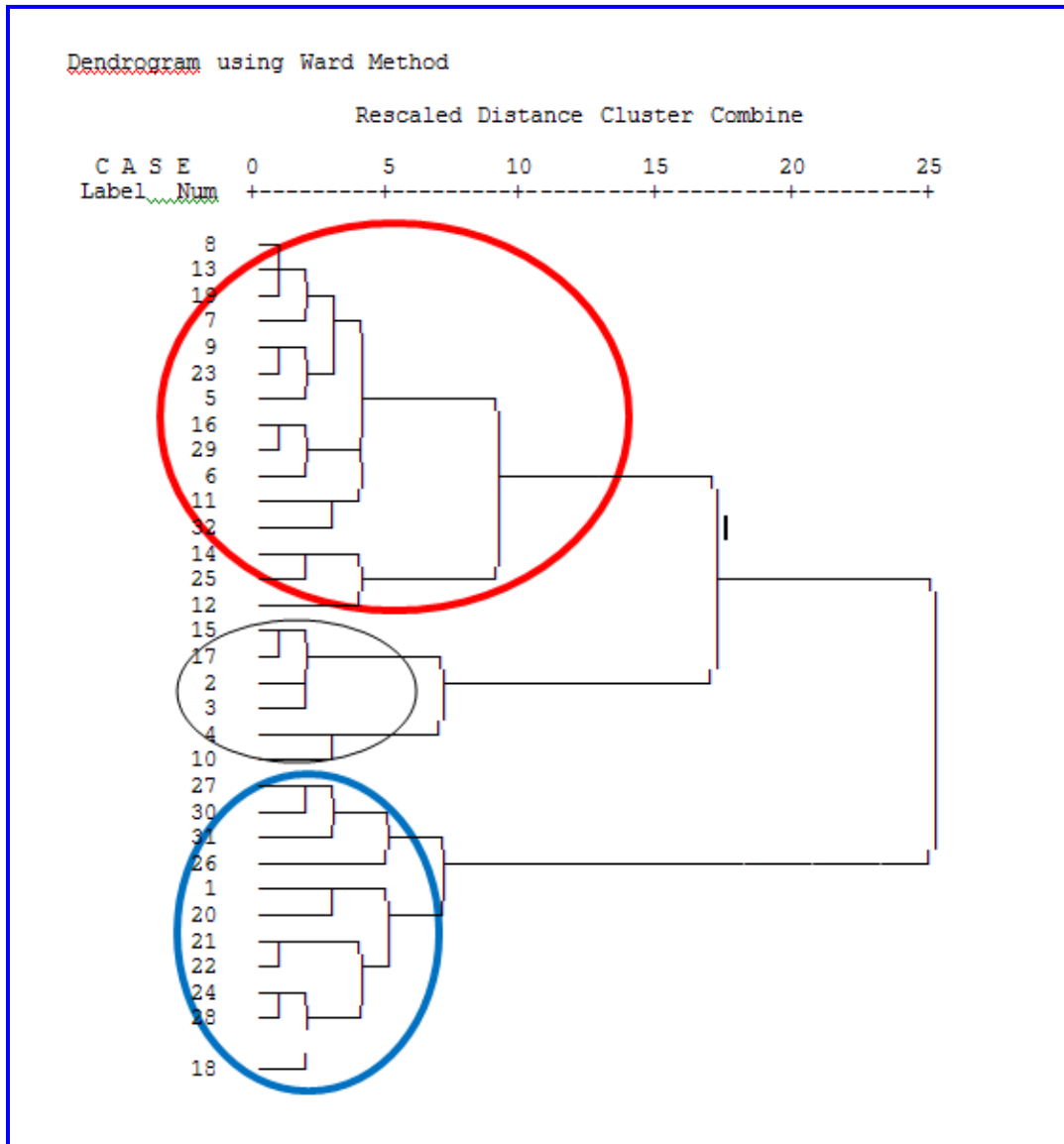
ภาพของ Dendrogram ในกรณีนี้มีแสดงในหน้าถัดไป



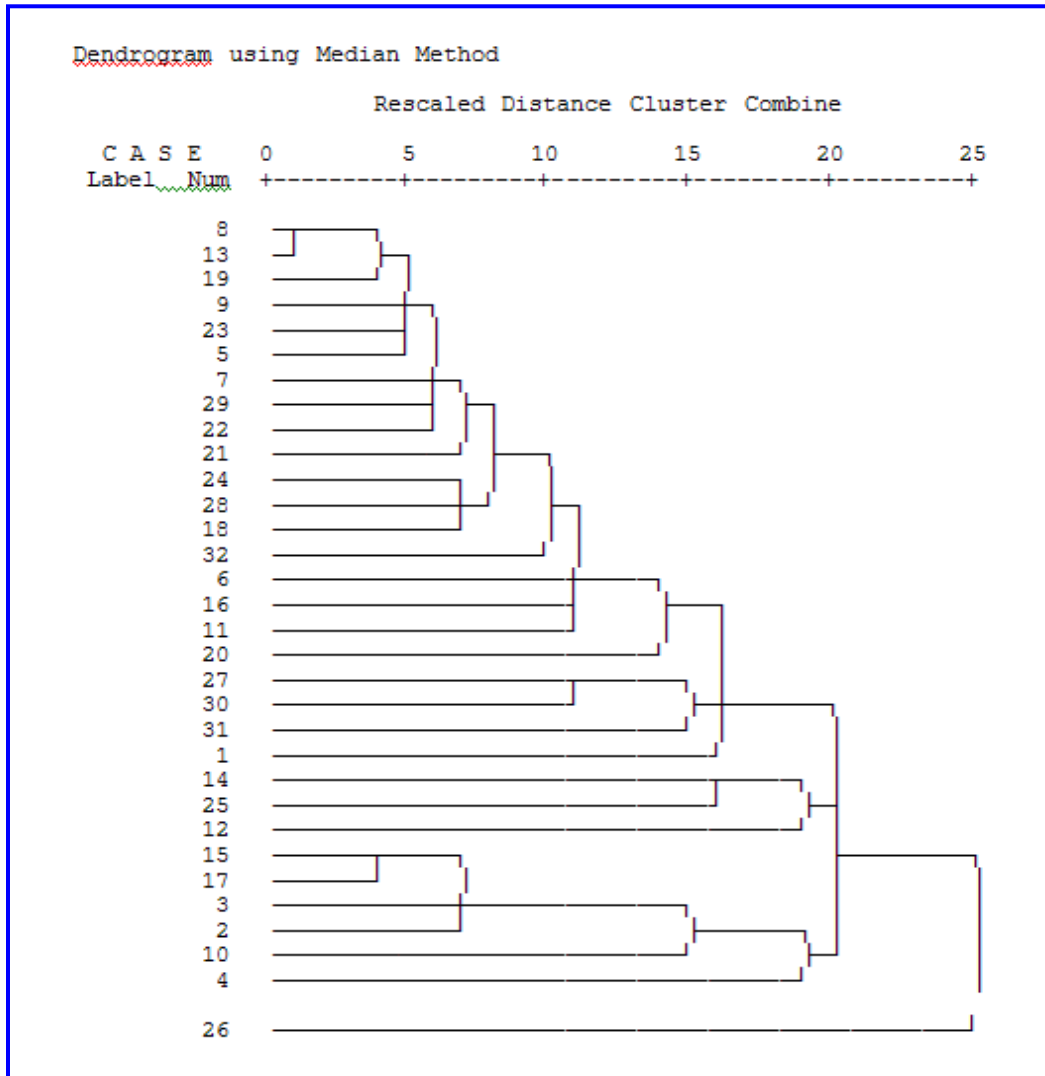
Hierarchical clustering โดยใช้ Ward's method

Dendrogram-Ward's method มีปรากฏให้เห็นในหน้าถัดไปซึ่งให้การยืนยันว่า 3-cluster solution

น่าจะเป็น solution ที่เหมาะสมที่สุด



ในขณะที่ dendrogram-median method แสดงได้ดังภาพต่อไปนี้



สิ่งหนึ่งที่นักวิเคราะห์สามารถสังเกตได้ก็คือ แม้การรวมกลุ่มหลายวิธีอาจจะชี้ให้เห็นตรงกันว่า 3-solution cluster เป็น solution ที่เหมาะสม แต่จะเห็นได้ว่า สมาชิกที่เคยอยู่รวมกับสมาชิกอื่นในclusterหนึ่งๆ โดยวิธีการรวมกลุ่มอย่างหนึ่ง อาจไม่ได้อยู่รวมกันอีกในcluster เดียวกันหากใช้วิธีการรวมกลุ่มที่แตกต่างออกไป

Contribution this issue: ดร.दनัย ปัตตพงศ์

อยากเรียนรู้การนำสถิติข้างต้นนี้ไปใช้ในการวิจัยระดับสารนิพนธ์ (independent study) วิทยานิพนธ์ (thesis) ดุษฎีนิพนธ์(dissertation) ปรึกษาได้ที่ [dpattaphongse@gmail.com](mailto:dpattaphongse@gmail.com)

- \* ผู้แต่ง MBA's Made Easy (160+ issues) เอกสารวิชาการด้านศาสตร์การบริหารธุรกิจที่ช่วยให้ธุรกิจสามารถยืนหยัดและอยู่รอดได้ในภาวะที่โลกเปลี่ยนแปลงอยู่ตลอดเวลา
- \* ผู้พัฒนา FINALYSIS... a dedicated software สำหรับให้บริการนักธุรกิจที่ต้องการวิเคราะห์ความเป็นไปได้ทางการเงินของโครงการพัฒนาอสังหาริมทรัพย์ (บ้านจัดสรร/จัดสรรที่ดินเพื่อการอุตสาหกรรม/อาคารชุด/อาคารสำนักงานให้เช่า) โรงแรม โรงพยาบาลเอกชน ห้างสรรพสินค้า โรงงานน้ำตาล โรงงานกระดาษ โรงไฟฟ้าชีวมวล ฯลฯ ได้เห็นตัวเลขก่อนโครงการเกิด หลีกเลี่ยงความผิดพลาดเป็นร้อยเป็นพันล้านหากเกิดการลงทุนจริง(กำหนด DEBUT 1 เมษายน 2569)
- \* ผู้แต่งหนังสือ”การวิเคราะห์ความเป็นไปได้ทางการเงินและการจัดวงเงินเครดิตของโครงการลงทุน”ประกอบด้วยตัวอย่างของธุรกิจจริงที่ไม่เปิดเผยชื่อนับ 100 บริษัท ครอบคลุมอุตสาหกรรม 24 อุตสาหกรรม
- \* Co-developer ซอฟต์แวร์ en@gex@cel<sup>®</sup> สำหรับใช้ทดสอบ/เรียนรู้ศัพท์(ประกอบด้วยแบบฝึกหัดและเฉลยกว่า 90 บทครอบคลุมศัพท์ระดับ SAT/IELTS/TOEFL กว่า 12,000 คำ) และไวยากรณ์อังกฤษ (ประกอบด้วยแบบฝึกหัดและเฉลยกว่า 160 บทหรือกว่า 10,000 ข้อครอบคลุมเนื้อหาระดับอุดมศึกษาและTOEFL) มาพร้อมกับไฟล์เสียง/ไฟล์ข้อมูล/ฯลฯ อีกมาก(กำหนด DEBUT 1 เมษายน 2569)