บทที่ 4
การจำแนกกลุ่มตัวแปรด้วย
เทคนิค Cluster Analysis

การจำแนกกลุ่มตัวแปรด้วยเทคนิค Cluster Analysis เป็นเทคนิคที่ใช้ในการจำแนกกลุ่มโดยไม่ทราบมาก่อนว่าควรมีกี่กลุ่ม แต่จะแบ่งตามค่าของตัวแปรที่นำมาใช้ในการแบ่ง โดยให้หน่วยที่อยู่ในกลุ่มเดียวกัน มีความคล้ายกันในตัวแปรที่ศึกษา แต่หน่วยที่อยู่ต่างกลุ่มกันจะมีความต่างกัน ดังนั้น การพิจารณาเลือกลักษณะหรือตัวแปรที่จะนำมาใช้ในการแบ่งกลุ่ม Case จึงมีความสำคัญ

จุดมุ่งหมายของการจำแนกกลุ่มตัวแปรด้วยเทคนิค Cluster Analysis

เพื่อเป็นการลดจำนวนโครงสร้างข้อมูล

เพื่อพัฒนาสมมติฐาน

เพื่อจำแนกกลุ่ม เช่น การแบ่งกลุ่มการใช้ประโยชน์ที่ดิน การทำความเข้าใจพฤติกรรมของประชาชนกลุ่มต่าง ๆ ในเขตเมือง หรือในเขตชนบท

ประเภทของเทคนิค Cluster Analysis

เทคนิค Cluster Analysis แบ่งเป็นหลายประเภทหรือเทคนิคย่อย โดยเทคนิคที่ใช้กันมากมี 2 เทคนิค คือ

1. เทคนิค Hierarchical Cluster Analysis

เป็นเทคนิคที่นิยมใช้กันมากในการแบ่งกลุ่ม Case หรือแบ่งกลุ่มตัวแปร โดยมีเงื่อนไขดังนี้

1. ในกรณีที่ใช้ในการแบ่ง Case นั้น จำนวน Case ต้องไม่มากนัก (จำนวน Case ควรต่ำกว่า 200 ถ้าตั้งแต่ 200 ขึ้นไปใช้ K-Means Cluster) และจำนวนตัวแปรต้องไม่มากเช่นกัน

2. ไม่จำเป็นต้องทราบจำนวนกลุ่มมาก่อน

3. ไม่จำเป็นเป็นต้องทราบว่าตัวแปรใดหรือ Case ใดอยู่กลุ่มใดก่อน

2. เทคนิค K-Means Cluster Analysis

เป็นเทคนิคการจำแนก Case ออกเป็นกลุ่มย่อย จะใช้เมื่อมีจำนวน Case มาก โดยจะต้องกำหนดจำนวนกลุ่ม หรือจำนวนCluster ที่ต้องการ เช่น กำหนดให้มี k กลุ่ม เทคนิค K – Mean จะมีการทำงานหลาย ๆ รอบ (Iteration) โดยในแต่ละรอบจะมีการรวม Cases ให้ไปอยู่ในกลุ่มใดกลุ่มหนึ่ง

โดยเลือกกลุ่มที่ Case นั้นมีระยะห่างจากค่ากลางของกลุ่มน้อยที่สุด แล้วคำนวณค่ากลางของกลุ่มใหม่ จะทำเช่นนี้จนกระทั่งค่ากลางของกลุ่มไม่เปลี่ยนแปลง หรือครบจำนวนรอบที่กำหนดไว้ สำหรับชนิดของตัวแปรที่ใช้ในเทคนิค K-Means Clustering จะต้องเป็นตัวแปรเชิงปริมาณ คือเป็นสเกลอันตรภาค (Interval Scale) หรือสเกลอัตราส่วน (Ration Scale) โดยไม่สามารถใช้กับข้อมูลที่อยู่ในรูปความถี่ หรือ Binary เหมือนเทคนิค Hierarchical

สำหรับข้อแตกต่างระหว่างเทคนิค Hierarchical กับวิธี K-Means มีดังนี้คือ

1. เทคนิค K-Means ใช้เมื่อมีจำนวน Case หรือจำนวนข้อมูลมาก โดยทั่วไปนิยมใช้เมื่อ $n \geq 200$ เพราะเมื่อ n มาก เทคนิค K-Means จะง่ายกว่า และใช้ระยะเวลาในการคำนวณน้อยกว่าการใช้เทคนิค Hierarchical หรือกล่าวได้ว่าเมื่อมีจำนวน Case ไม่มากควรใช้เทคนิค Hierarchical

2. เทคนิค K-Means นั้น ผู้ใช้จะต้องกำหนดจำนวนกลุ่มที่แน่นอนไว้ล่วงหน้า

3. เทคนิค Hierarchical นั้น ผู้วิเคราะห์จะ Standardized ข้อมูลหรือไม่ก็ได้ แต่โดยวิธี K-Means จะต้องทำการ Standardized ข้อมูลก่อนเสมอ

ตัวอย่างงานที่ใช้การวิเคราะห์จำแนกกลุ่ม

- ใช้แบ่งพื้นที่เสี่ยงภัยแล้ง โดยอาจจะพิจารณาจากตัวแปรดัชนีพืชพรรณ (Normalized Difference Vegetation Index : NDVI), ดัชนีความแตกต่างความชื้น (Normalized Difference Water Index : NDWI), ดัชนีเงื่อนไขอุณหภูมิ (Temperature Condition Index : TCI) ดัชนีเงื่อนไขพืชพรรณ (The Vegetation Condition Index : VCI), ดัชนีความชื้นในดิน (Soil Moisture Index) เป็นต้น

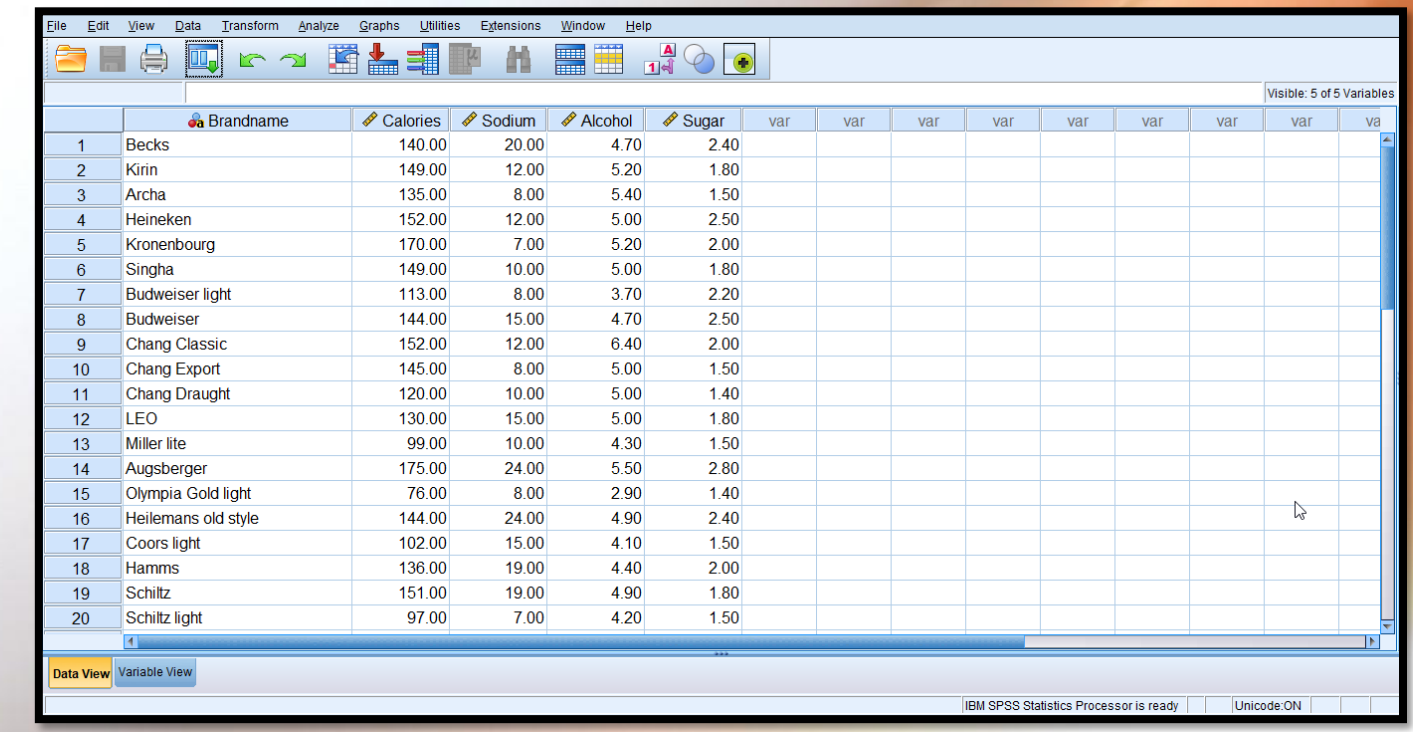
- ใช้แบ่งกลุ่มประเทศเป็นกลุ่มประเทศพัฒนาแล้ว กลุ่มประเทศกำลังพัฒนา และกลุ่มประเทศด้อยพัฒนา โดยอาจจะพิจารณาจากตัวแปรลักษณะทางเศรษฐกิจ ดัชนีการพัฒนามนุษย์ การส่งเสริมคุณภาพชีวิตและสาธารณสุข รายได้ประชากร เป็นต้น

- ใช้ศึกษา การแบ่งส่วนตลาด ซึ่งจะทำให้สามารถวางกลยุทธ์ทางการตลาดได้อย่างมีประสิทธิภาพมากขึ้น การที่จะสามารถแยกกลุ่มผู้บริโภคออกเป็นกลุ่มย่อยได้ จะต้องพิจารณาถึงตัวแปรที่ใช้ในการแบ่งกลุ่มผู้บริโภค ที่จะทำให้ผู้ที่อยู่ต่างกลุ่มกันมีพฤติกรรมการบริโภคที่แตกต่างกัน ตัวแปรดังกล่าวอาจจะประกอบด้วยเกณฑ์ด้านประชากรศาสตร์ (Demographic) เกณฑ์ด้านจิตวิทยา (Psychographic) เกณฑ์ด้านพฤติกรรม (Behavioral) และเกณฑ์ด้านปัจจัยทางภูมิศาสตร์ (Geographic) เป็นต้น

ตัวอย่างการจำแนกกลุ่มตัวแปรด้วยเทคนิค Cluster Analysis โดยใช้โปรแกรมโปรแกรมสำเร็จรูปทางสถิติ

ตัวอย่างการใช้เทคนิค Hierarchical Cluster Analysis

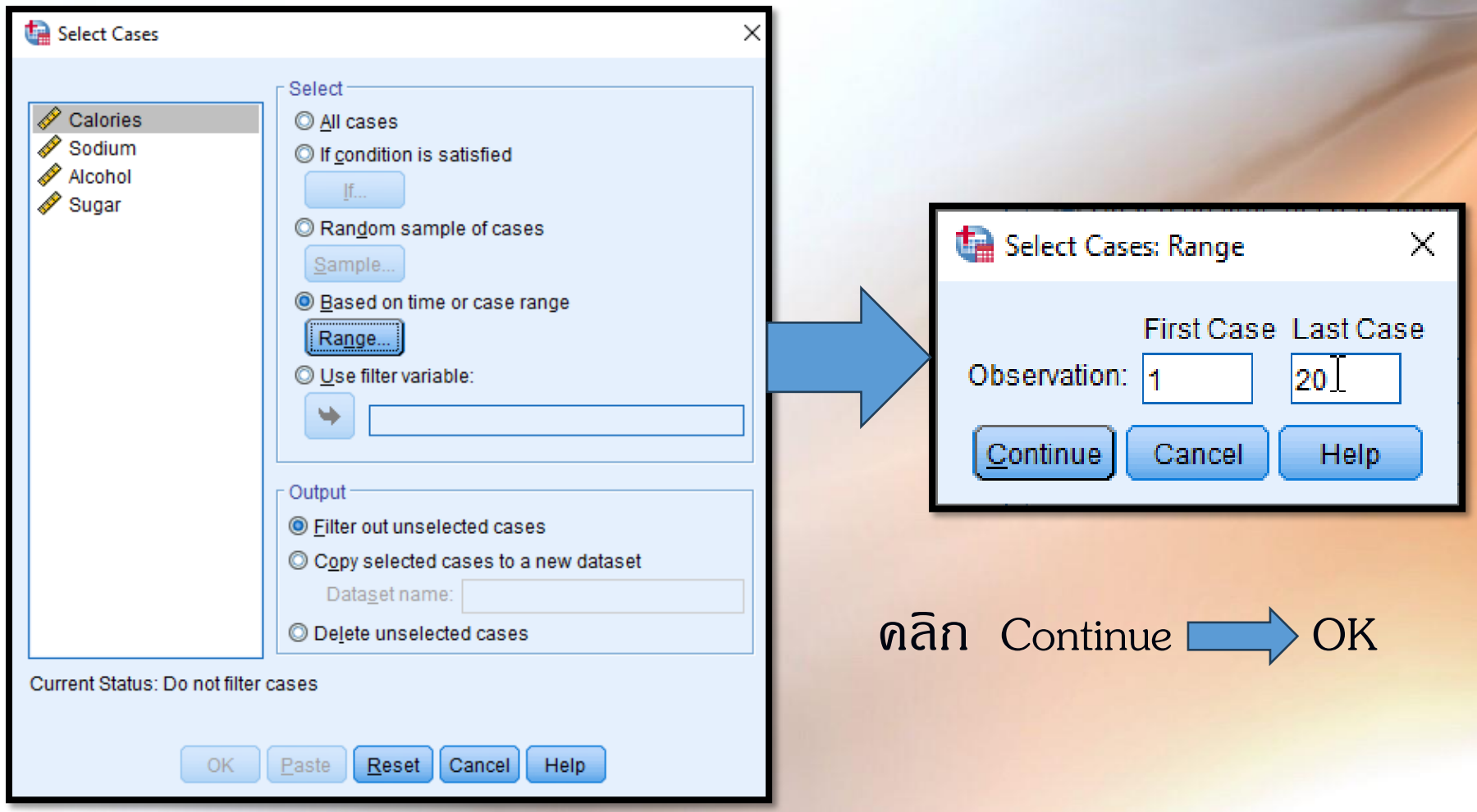
ตัวอย่าง จาก File clusteryuth1_Ex4.1 ผู้ศึกษาสนใจที่จะจัดกลุ่มเบียร์ 20 ยี่ห้อว่าจะรวมกันอย่างไร โดยคาดว่าตัวแปร แคลอรี่ โซเดียม แอลกอฮอล์ และ น้ำตาลมีผลต่อการรวมกลุ่ม ดังภาพประกอบ



	Brandname	Calories	Sodium	Alcohol	Sugar
1	Becks	140.00	20.00	4.70	2.40
2	Kirin	149.00	12.00	5.20	1.80
3	Archa	135.00	8.00	5.40	1.50
4	Heineken	152.00	12.00	5.00	2.50
5	Kronenbourg	170.00	7.00	5.20	2.00
6	Singha	149.00	10.00	5.00	1.80
7	Budweiser light	113.00	8.00	3.70	2.20
8	Budweiser	144.00	15.00	4.70	2.50
9	Chang Classic	152.00	12.00	6.40	2.00
10	Chang Export	145.00	8.00	5.00	1.50
11	Chang Draught	120.00	10.00	5.00	1.40
12	LEO	130.00	15.00	5.00	1.80
13	Miller lite	99.00	10.00	4.30	1.50
14	Augsberger	175.00	24.00	5.50	2.80
15	Olympia Gold light	76.00	8.00	2.90	1.40
16	Heilemans old style	144.00	24.00	4.90	2.40
17	Coors light	102.00	15.00	4.10	1.50
18	Hamms	136.00	19.00	4.40	2.00
19	Schiltz	151.00	19.00	4.90	1.80
20	Schiltz light	97.00	7.00	4.20	1.50

ขั้นตอนที่ 1 Click Data ➡ Select Cases ... Click Based on time or case range

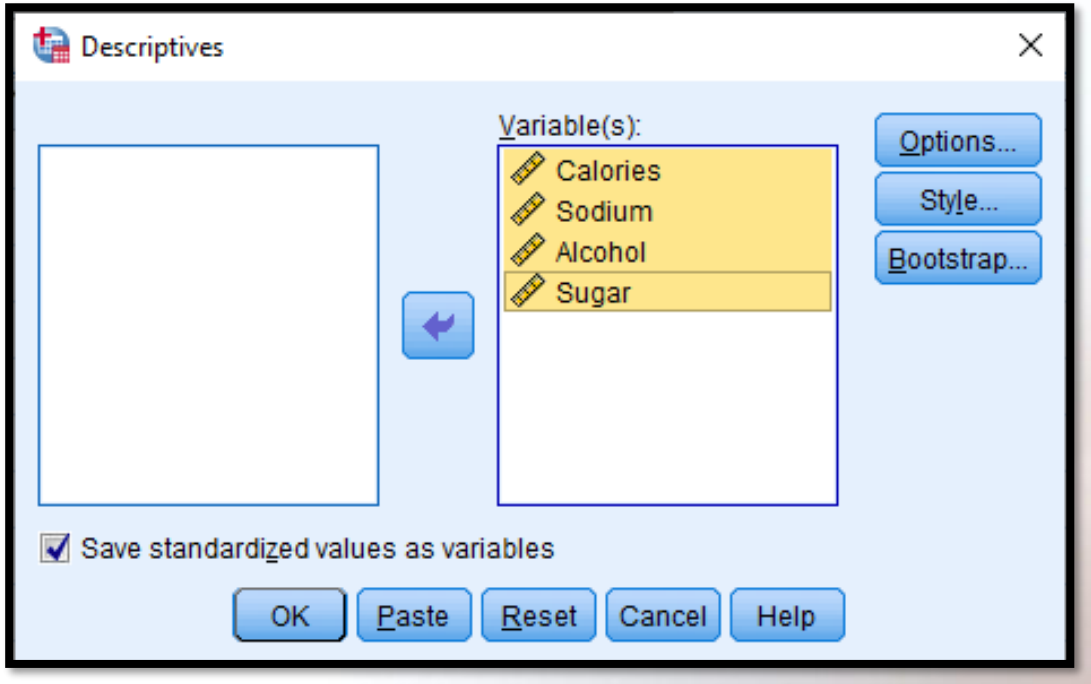
จากนั้น Click range ใส่เลข 1 ที่ First Case และเลข 20 ที่ Last Case จะได้หน้าจอ ดังภาพประกอบ



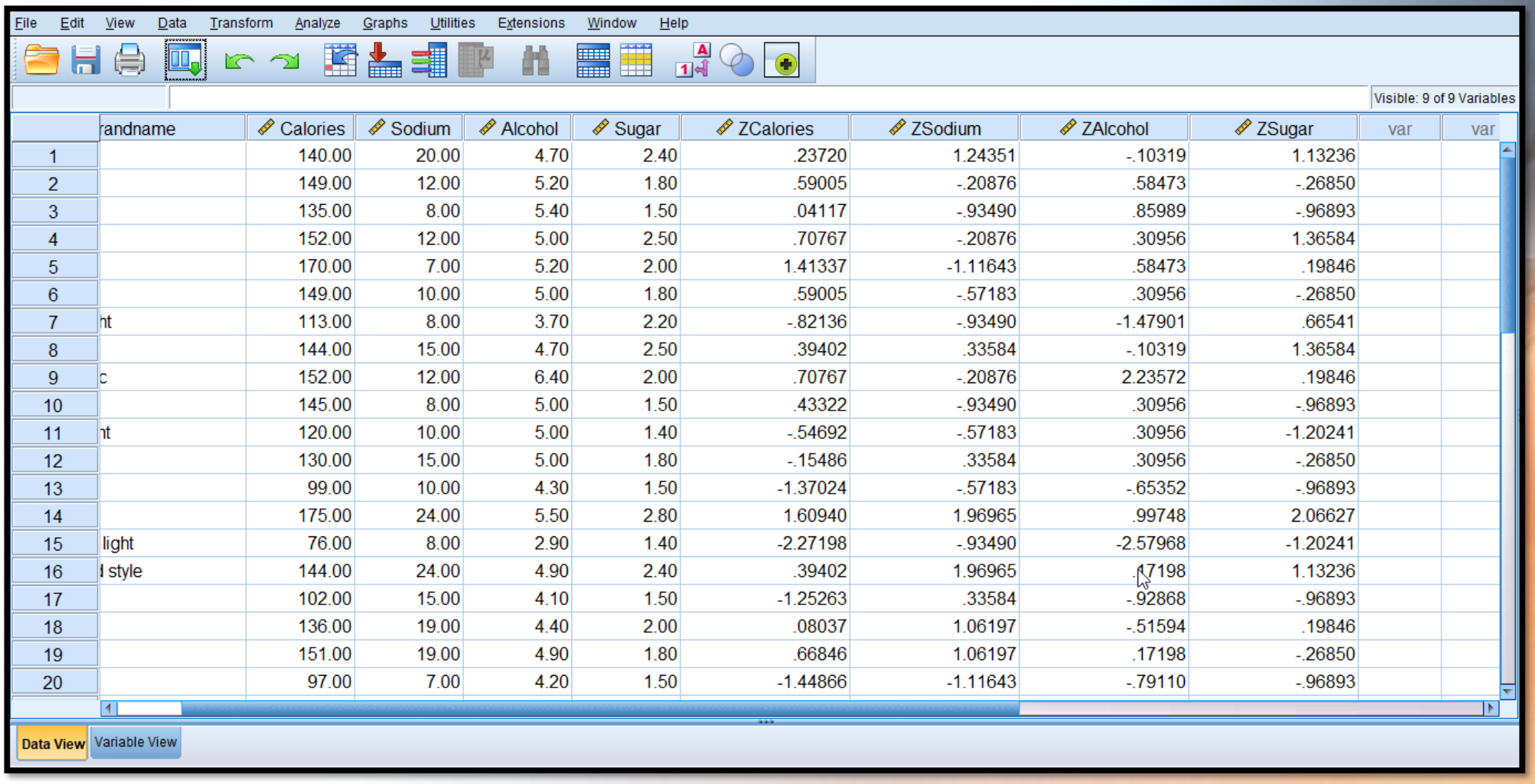
คลิก Continue ➡ OK

ขั้นตอนที่ 2 แปลงตัวแปรให้เป็นมาตรฐาน โดยให้ Standardized ข้อมูลให้เป็น Z – Score ของทุกตัวแปรที่นำมาจัดกลุ่มสมาชิก โดยใช้คำสั่ง ดังนี้

1) คลิกที่ Analyze
2) เลือก Descriptive Statistics
3) คลิกที่ Descriptive จะได้หน้าจอ ดังภาพประกอบ



ในหน้าจอ Descriptive เลือกตัวแปรอย่างน้อย 1 ตัวแปร ใส่ในช่อง Variable(s) ในที่นี้เลือกตัวแปรทุกตัวแปรที่จะนำมาจัดกลุ่มสมาชิก จากนั้นให้คลิก Save standardized value as variables ตัวแปรที่แปลงให้เป็นมาตรฐานแล้วจะอยู่ในรูป Z – Score โดยจะอยู่ในแฟ้มข้อมูลเดิม ดังภาพประกอบ



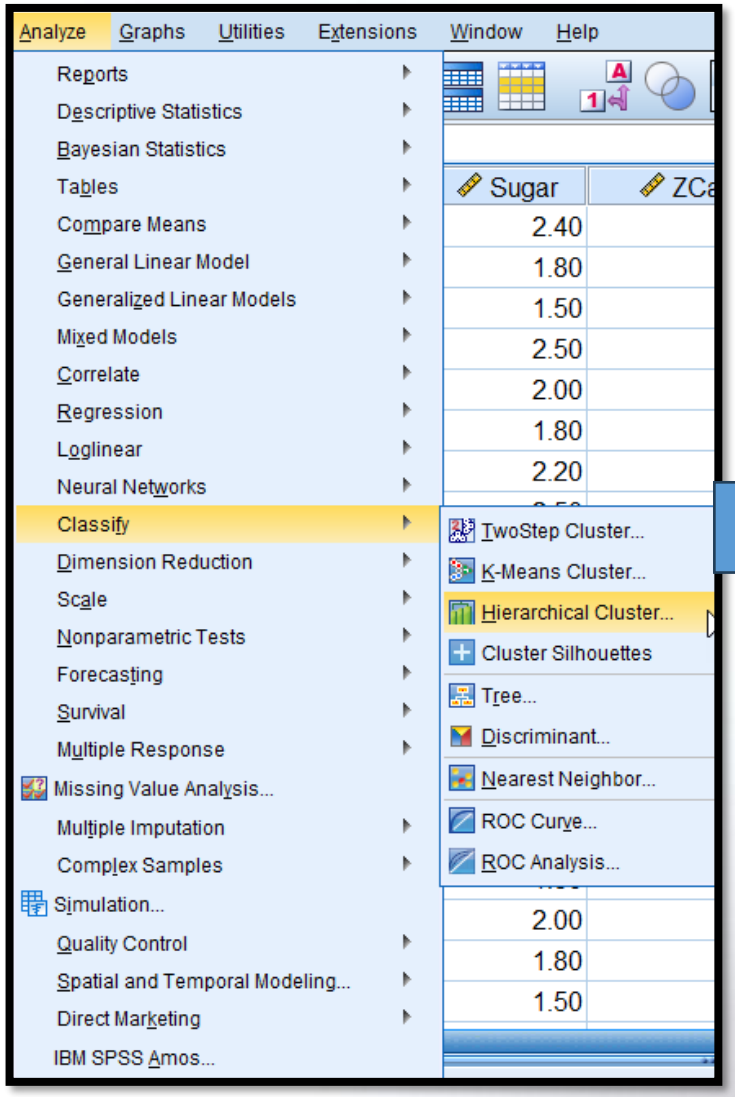
File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help

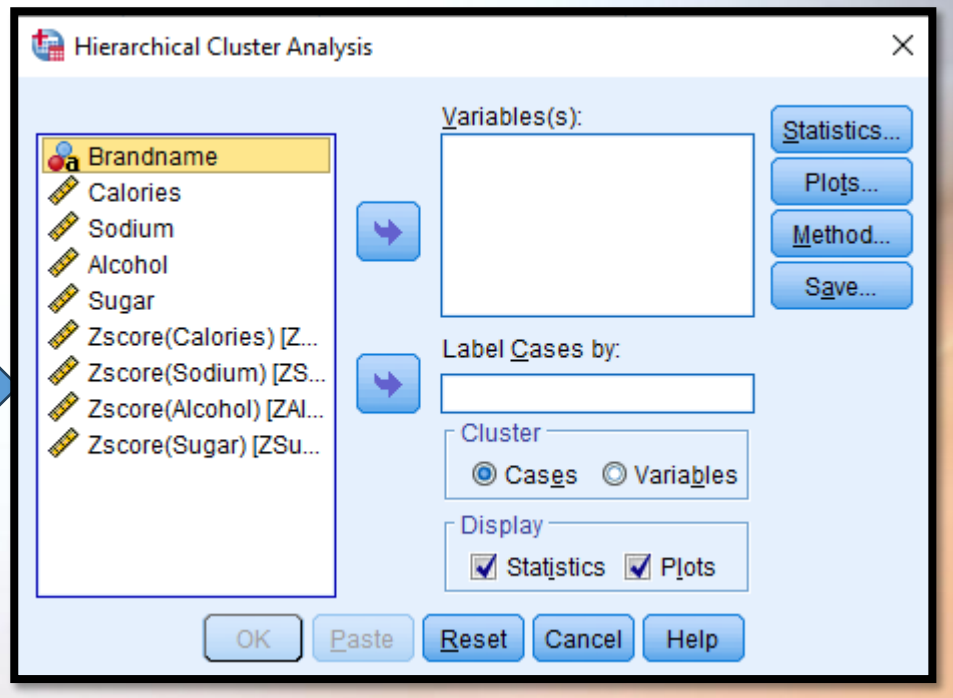
Visible: 9 of 9 Variables

	randname	Calories	Sodium	Alcohol	Sugar	ZCalories	ZSodium	ZAlcohol	ZSugar	var	var
1		140.00	20.00	4.70	2.40	.23720	1.24351	-.10319	1.13236		
2		149.00	12.00	5.20	1.80	.59005	-.20876	.58473	-.26850		
3		135.00	8.00	5.40	1.50	.04117	-.93490	.85989	-.96893		
4		152.00	12.00	5.00	2.50	.70767	-.20876	.30956	1.36584		
5		170.00	7.00	5.20	2.00	1.41337	-1.11643	.58473	.19846		
6		149.00	10.00	5.00	1.80	.59005	-.57183	.30956	-.26850		
7	nt	113.00	8.00	3.70	2.20	-.82136	-.93490	-1.47901	.66541		
8		144.00	15.00	4.70	2.50	.39402	.33584	-.10319	1.36584		
9	c	152.00	12.00	6.40	2.00	.70767	-.20876	2.23572	.19846		
10		145.00	8.00	5.00	1.50	.43322	-.93490	.30956	-.96893		
11	nt	120.00	10.00	5.00	1.40	-.54692	-.57183	.30956	-1.20241		
12		130.00	15.00	5.00	1.80	-.15486	.33584	.30956	-.26850		
13		99.00	10.00	4.30	1.50	-1.37024	-.57183	-.65352	-.96893		
14		175.00	24.00	5.50	2.80	1.60940	1.96965	.99748	2.06627		
15	light	76.00	8.00	2.90	1.40	-2.27198	-.93490	-2.57968	-1.20241		
16	l style	144.00	24.00	4.90	2.40	.39402	1.96965	.17198	1.13236		
17		102.00	15.00	4.10	1.50	-1.25263	.33584	-.92868	-.96893		
18		136.00	19.00	4.40	2.00	.08037	1.06197	-.51594	.19846		
19		151.00	19.00	4.90	1.80	.66846	1.06197	.17198	-.26850		
20		97.00	7.00	4.20	1.50	-1.44866	-1.11643	-.79110	-.96893		

Data View Variable View

ขั้นตอนที่ 3 จัดกลุ่มสมาชิกโดยใช้คำสั่ง Analyze ➡ Classify ➡ Hierarchical Cluster จะได้หน้าจอ ดังภาพประกอบ





หน้าจอ Hierarchical Cluster Analysis มีคำสั่งให้เลือก ดังนี้

1) ส่วน Variable (s) : เป็นคำสั่งให้เลือกว่า ถ้าต้องการจัดกลุ่มสมาชิกให้เลือกตัวแปรที่เป็นตัวเลข (Numeric Variable) มาอย่างน้อย 1 ตัวแปร และถ้าต้องการจัดกลุ่มตัวแปรให้เลือกตัวแปรที่เป็นตัวเลขมาอย่างน้อย 3 ตัวแปร

2) ส่วน Label Cases By : เป็นคำสั่งให้นำชื่อสมาชิกที่ต้องการจัดกลุ่มมาใส่ช่องนี้ถ้าเป็นการจัดกลุ่มสมาชิก ในที่นี้คือ ยี่ห้อเบียร์ (Brandname) ให้นำตัวแปร Brandname มาใส่ โดยที่ตัวแปรช่องนี้จะมีการวัดเป็นตัวแปร Normal และชนิดของตัวแปรเป็น String

3) ส่วน Cluster มี 2 ทางเลือก คือ

3.1 เลือก Cases กรณีจัดกลุ่มสมาชิก

3.2 เลือก Variable กรณีจัดกลุ่มตัวแปร

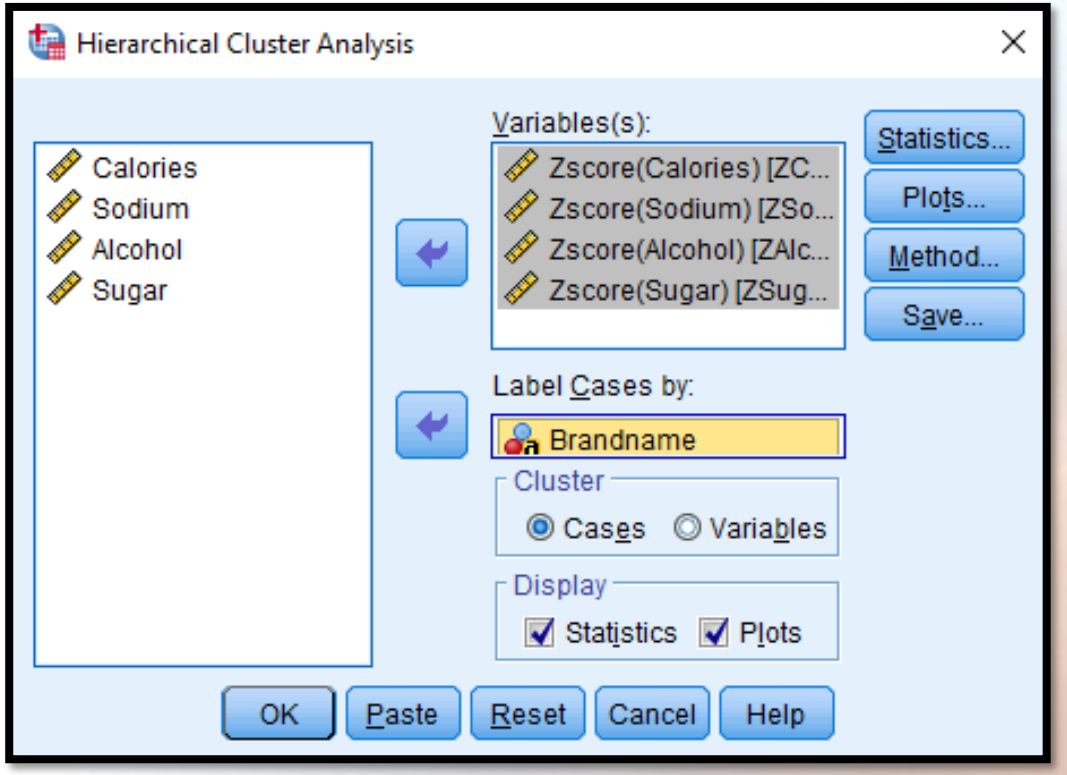
ในที่นี้เป็นการจัดกลุ่มสมาชิก เลือก Cases ตามโปรแกรม

4) ส่วน display : เป็นคำสั่งเลือกผลลัพธ์จากโปรแกรม ดังนี้

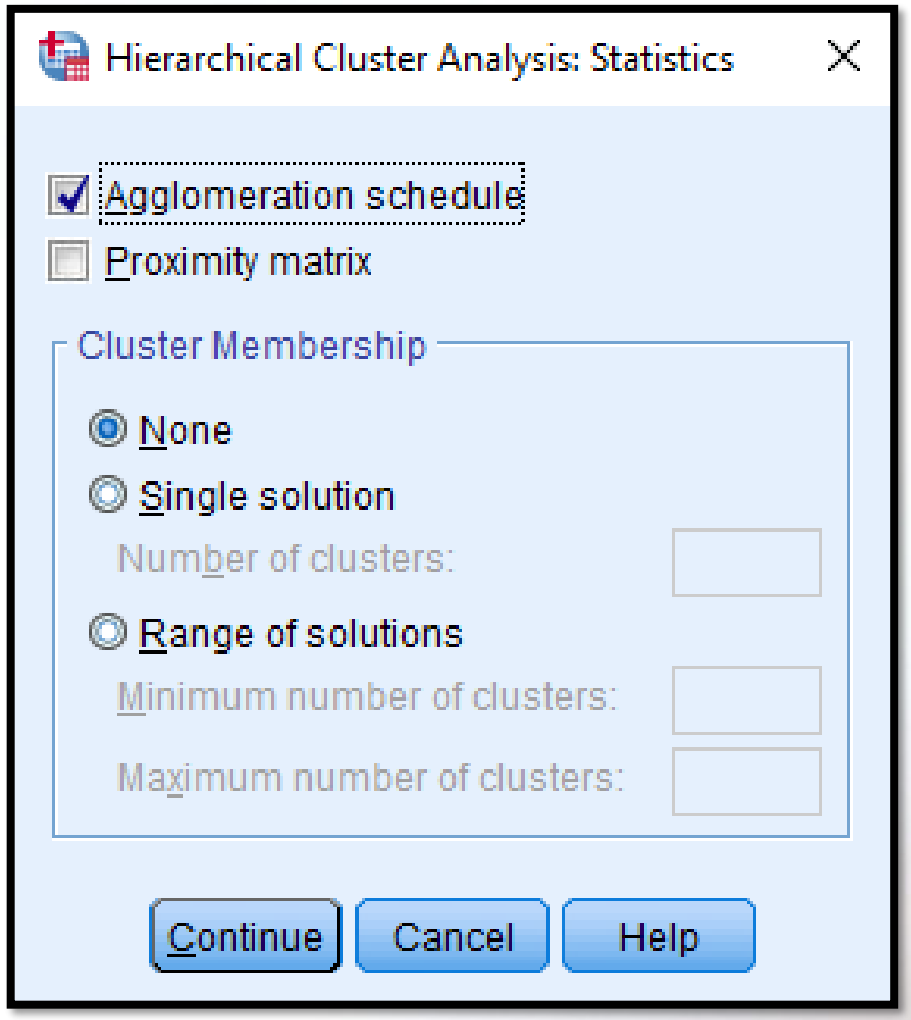
☑ Statistics ให้แสดงค่าสถิติ

☑ Plots ให้แสดงกราฟ

จะได้ดังภาพประกอบ



คลิก Statistics จะได้ Hierarchical Cluster Analysis : Statistics ซึ่งมีคำสั่งดังภาพประกอบ



1. Agglomeration schedule เป็นคำสั่งแสดงขั้นตอนการรวมกลุ่มสมาชิก

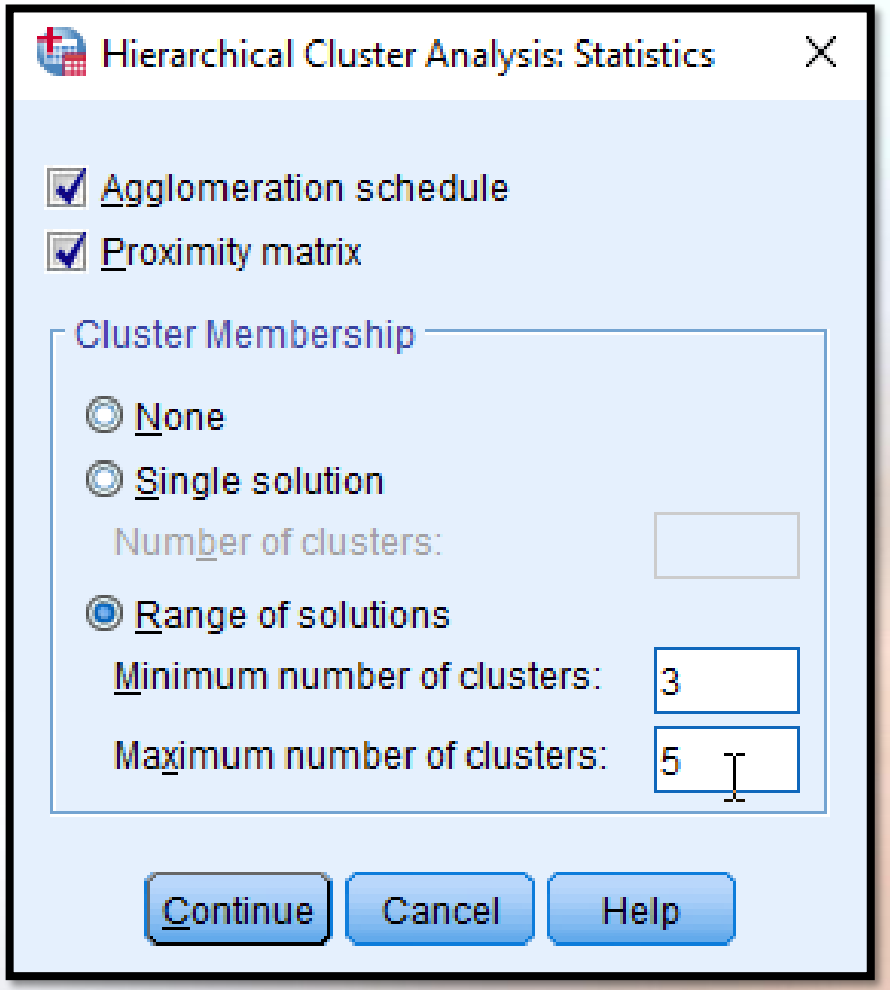
2. Proximity matrix เป็นคำสั่งให้แสดงระยะห่างของสมาชิกแต่ละคู่ โดยระยะห่างที่แสดงคือ ระยะห่างยูคลิเดียนกำลังสอง (Squared Euclidean Distance)

3. Cluster membership เป็นคำสั่งให้แสดงแต่ละสมาชิกอยู่ในกลุ่มใด ซึ่งผู้วิเคราะห์มีทางเลือก ดังนี้

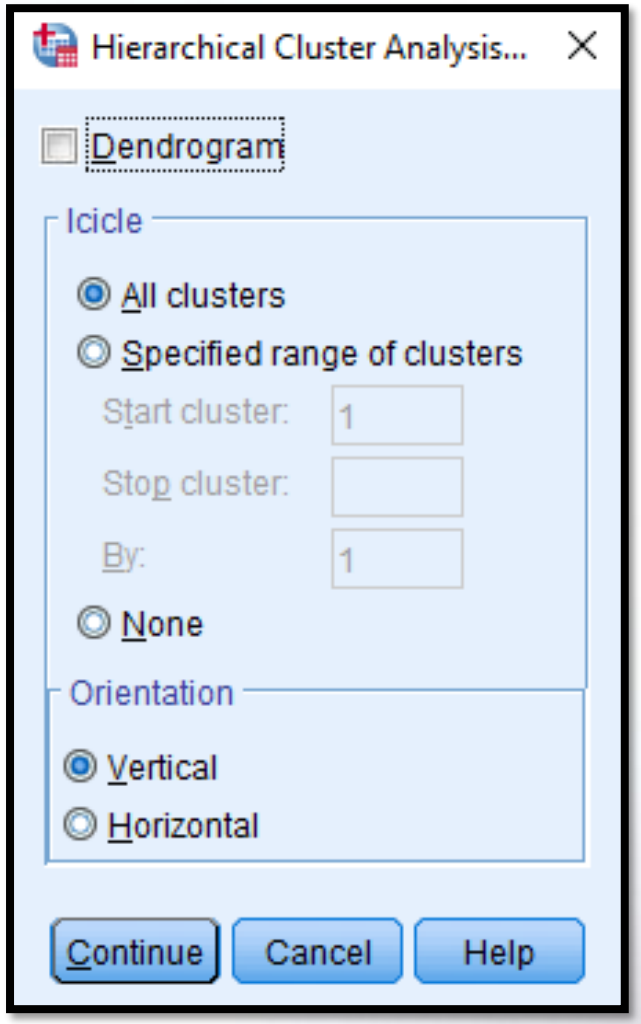
3.1 None : ไม่แสดงการเป็นสมาชิกของแต่ละสมาชิก

3.2 Single Solution : ให้แสดงสมาชิกของกลุ่ม (Cluster) โดยให้ผู้วิเคราะห์กำหนดจำนวนกลุ่ม (Number of Cluster) ที่เป็นเลขจำนวนเต็ม ทั้งนี้ควรกำหนดค่าตั้งแต่ 2 ขึ้นไป

4. Range of Solutions: เป็นคำสั่งให้แสดงจำนวนกลุ่มเป็นช่วงต่ำสุดและสูงสุดที่คิดว่าควรจะเป็น โดยค่าต่ำสุดต้องกำหนดเป็นค่า 2 อย่างต่ำ ในที่นี้กำหนดดังภาพประกอบ



คลิก Continue ➡ Plots... จะได้หน้าจอ Hierarchical Cluster Analysis : Plot ซึ่งมีคำสั่งดังภาพประกอบ



1. ที่ Dendrogram เป็นกราฟที่เป็นการรวมสมาชิกต่าง ๆ ให้อยู่ในกลุ่มเดียวกัน จากนั้นจึงพิจารณาว่าควรจะแบ่งสมาชิกออกเป็นกี่กลุ่ม (Cluster) จึงจะเหมาะสม ซึ่งในกราฟจะแสดงระยะห่างของการรวมกลุ่ม โดยมีระยะห่างอยู่ในช่วง 0 – 25

2. ที่ Icicle เป็นคำสั่ง Plots อีกแบบหนึ่งที่มีคำสั่งย่อยให้เลือก 3 ทาง ได้แก่

2.1 All cluster ให้แสดง Icicle Plot ทุกสมาชิก

2.2 Specified range of clusters เป็นคำสั่งให้แสดง Icicle Plot ตามช่วงของจำนวนกลุ่ม (Cluster) ที่ต้องการ โดยที่

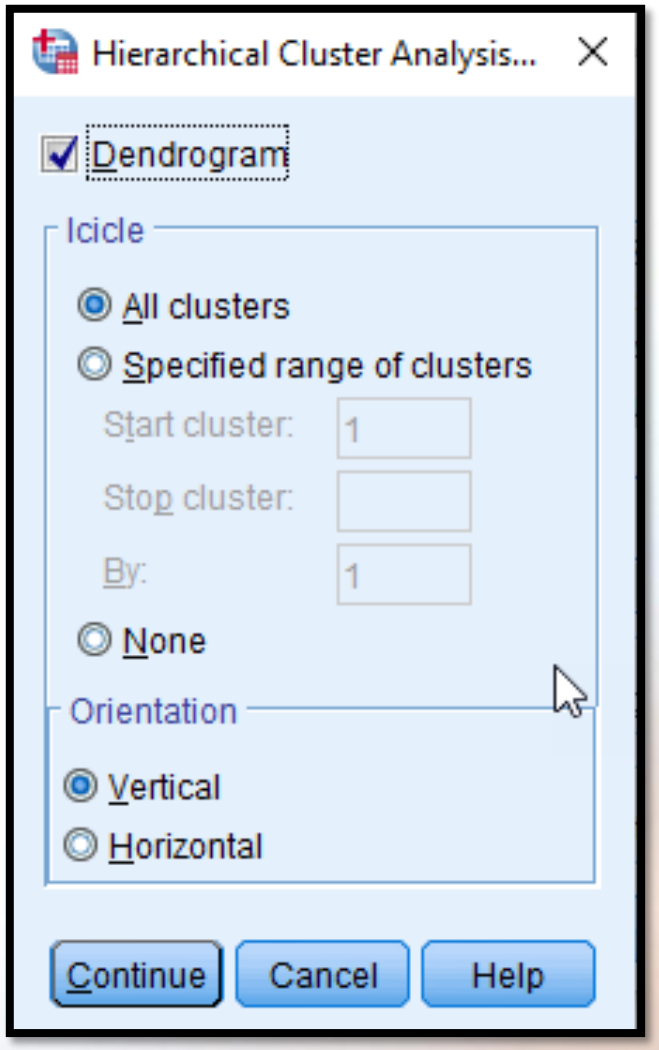
- Start Cluster ใส่กลุ่มเริ่มต้น
- Stop Cluster ใส่กลุ่มที่ต้องการสุดท้าย
- By หมายถึง การเพิ่มขึ้นครั้งละ เช่น ใส่เลข 5 จะทำให้ Icicle Plot แสดง 5 กลุ่ม

2.3 None เป็นคำสั่งไม่ให้แสดง Icicle Plot

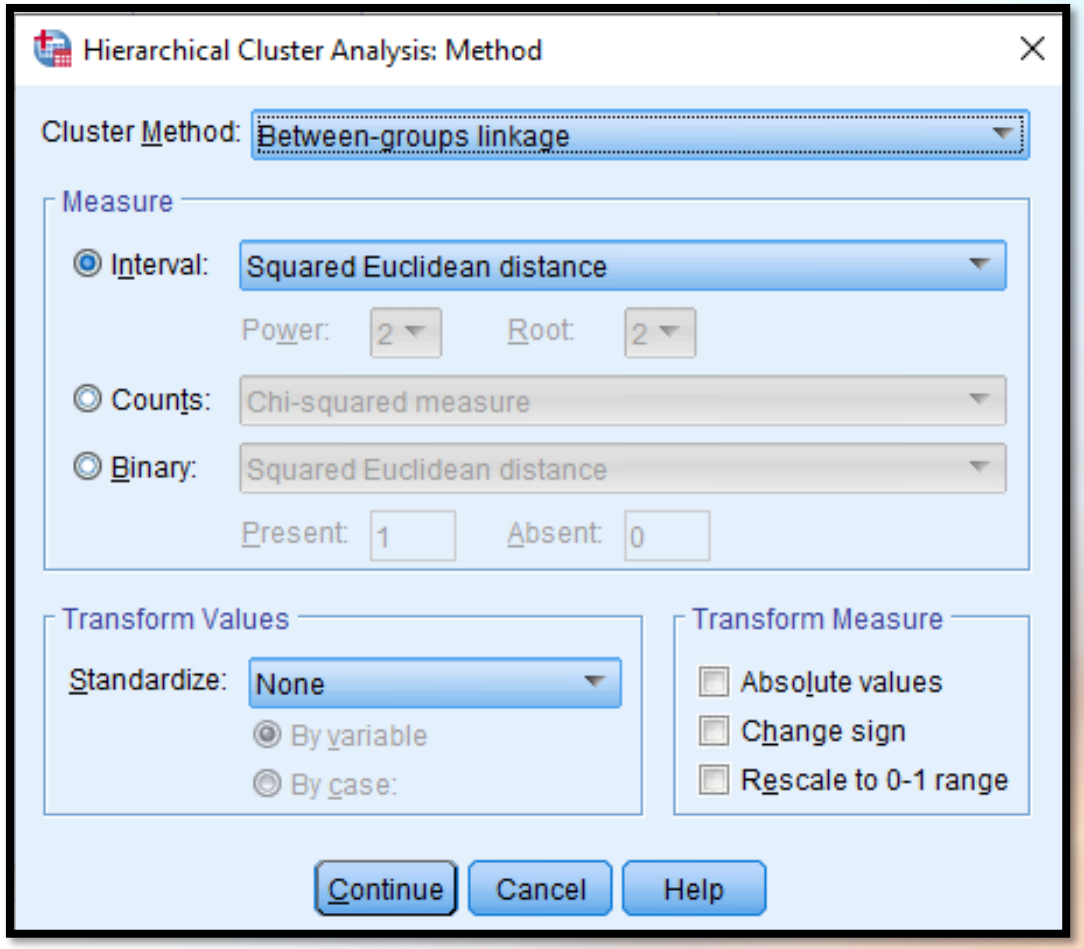
3. ที่ Orientation เป็นคำสั่งให้แสดง Icicle Plots 2 ทางเลือกคือ

3.1 Verticle แสดง Icicle Plots แนวตั้ง

3.2 Horizontal แสดง Icicle Plots แนวนอน เลือกคำสั่งต่าง ๆ ดังภาพประกอบ



คลิก Continue ➡ Method... จะได้หน้าจอดังภาพประกอบ



หน้าจอ Hierarchical Cluster Analysis : Method จะมีคำสั่งย่อยดังนี้

1. ที่ Cluster Method ให้เลือกวิธีการจัดกลุ่ม

2. ที่ Measure เป็นคำสั่งวัดระยะห่างและวัดความคล้าย ซึ่งการวัดทั้ง 2 อย่างจะขึ้นอยู่กับชนิดของข้อมูลที่กำหนดเป็น 3 ประเภท คือ

2.1 Interval ใช้กับข้อมูล Internal Scale และ Ratio Scale การวัดระยะห่างและความคล้าย

2.2 Count ใช้กับข้อมูลที่อยู่ในรูปความถี่

2.3 Binary ใช้กับข้อมูลที่มีค่าได้ 2 ค่า ซึ่งจะแสดงเป็นตาราง 2× 2 ของสมาชิกแต่ละคู่ในคำสั่งของโปรแกรม SPSS ที่ Present แสดงข้อมูลที่สร้างขึ้น มีค่าเป็น 1 หมายถึงสมาชิกที่มีลักษณะที่กำหนดไว้ และ 0 หมายถึง สมาชิกที่ไม่มีลักษณะตามที่กำหนด

3. Transform Values เป็นคำสั่งที่ต้องการเปลี่ยนค่าของสมาชิกหรือตัวแปรเพื่อให้ความสำคัญเท่ากัน เมื่อข้อมูลเดิมมีสเกลต่างกัน โดยจะกำหนด Standardize ข้อมูลและมีทางเลือกคือ

3.1 None ไม่ปรับเปลี่ยนข้อมูลเดิม

3.2 Z Score เปลี่ยนเป็นค่ามาตรฐานที่มีค่าเฉลี่ยเป็น 0 และค่า S.D. เป็น 1

3.3 Range-1 to 1 เปลี่ยนเป็นค่ามาตรฐานมีค่า -1 ถึง 1

3.4 Range 0 to 1 เปลี่ยนเป็นค่ามาตรฐาน มีค่า 0 ถึง 1

เมื่อกำหนดแล้วให้เลือกคำสั่งย่อยอีกคือ

(1) By Variable เปลี่ยนค่าตัวแปร

(2) By Case เปลี่ยนค่าตัวแปร

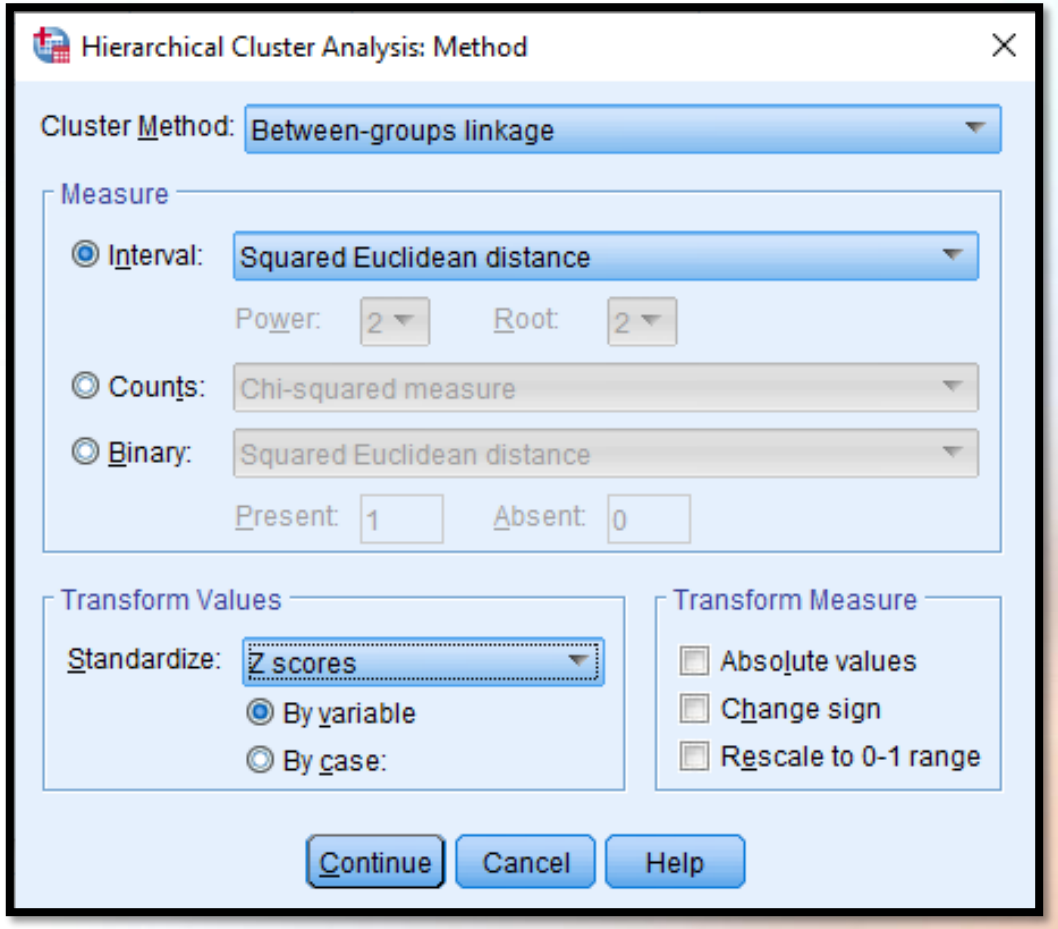
4. Transform Measure เมื่อใช้คำสั่ง Interval หรือคำสั่ง Counts แล้วในการ Standardize ข้อมูลสมาชิก โดยมีทางเลือกดังนี้

(1) Absolute Value เป็นการคำนวณค่าสัมบูรณ์ของระยะห่างระหว่างสมาชิก

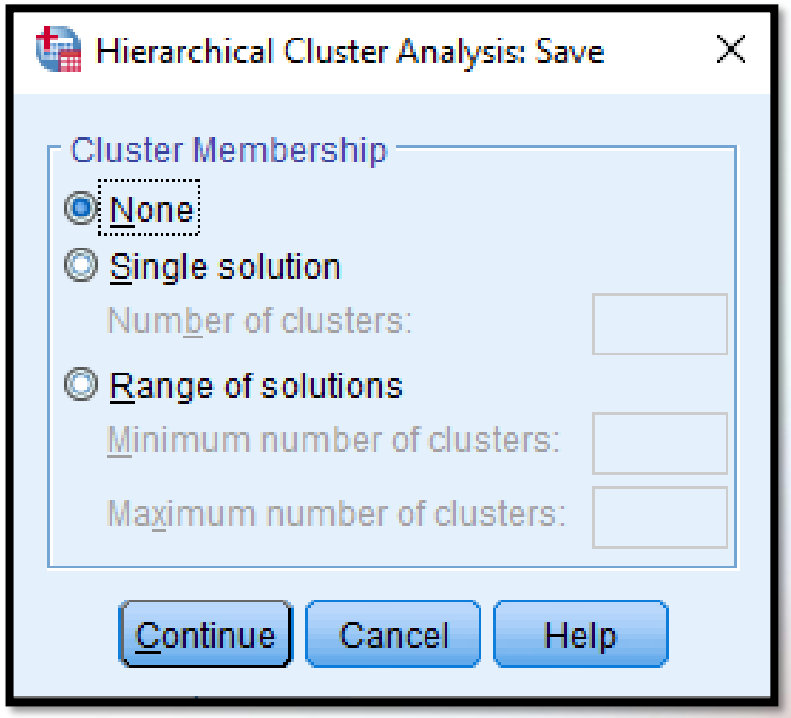
(2) Change Sign เป็นคำสั่งเปลี่ยนความไม่คล้ายเป็นความคล้ายหรือเปลี่ยนความคล้ายเป็นความไม่คล้ายของสมาชิก

(3) Rescale to 0 – 1 Range เปลี่ยนระยะห่างของสมาชิกให้มีค่า 0 ถึง 1 ซึ่งก็คือการทำ Standardize อย่างหนึ่ง

ในที่นี้เลือกตามภาพประกอบ ดังนี้



คลิก Continue ➡ Save จะได้หน้าจอ Hierarchical Cluster Analysis Save New Variable ดังนี้



หน้าจอ Save New Variable มีคำสั่งคือ

(1) None ไม่ต้องบันทึกเลขที่กลุ่ม

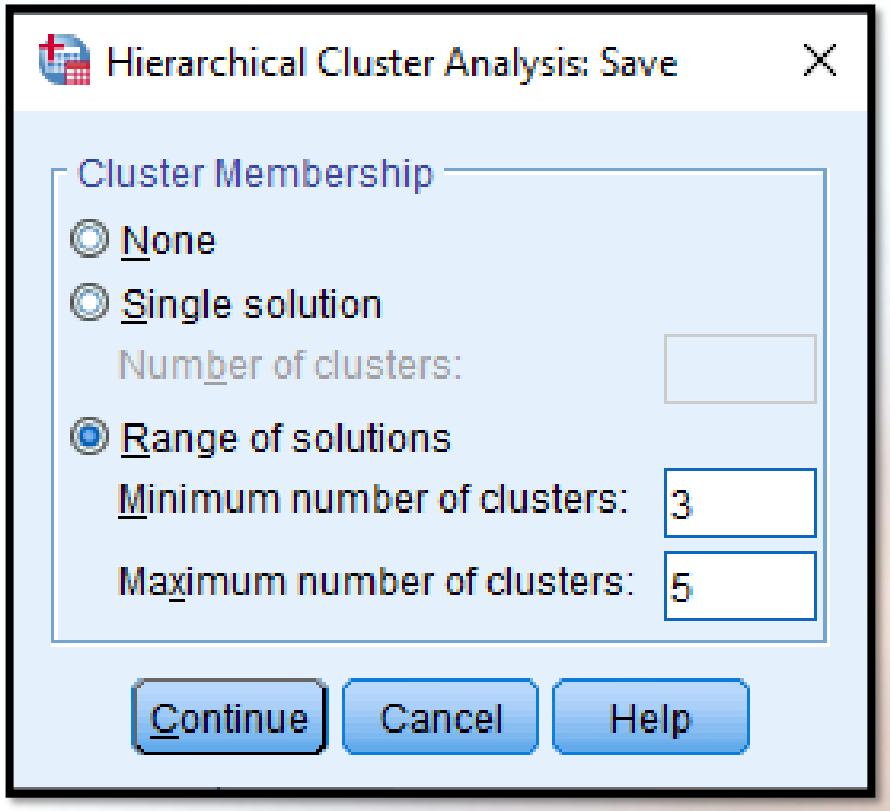
(2) Single Solution บันทึกกลุ่มที่แน่นอนค่าเดียว

(3) Range of solutions บันทึกเลขที่กลุ่มที่กำหนดว่ามีจำนวนกลุ่มหลาย ๆ แบบ เช่น ถ้าจำนวนบันทึกเลขที่กลุ่มของแต่ละสมาชิก เมื่อแบ่งเป็น 3, 4 และ 5

Minimum number of Cluster ใส่เลข 3

Maximum number of Cluster ใส่เลข 5

จะเลือกได้ดังภาพประกอบ



คลิก Continue ➡ OK โปรแกรมจะแสดงผลดังนี้

Proximity Matrix

Case	1:Becks	2:Kirin	3:Archa	4:Heineken	5: Kronenbourg	6:Singha	7:Budweiser light	8:Budweiser	9:Chang Classic	10:Chang Export	11:Chang Draught	12:LEO
	Squared Euclidean Distance											
1:Becks	.000	4.669	10.127	2.555	8.298	5.553	7.977	.903	8.673	9.370	9.532	3.11
2:Kirin	4.669	.000	1.395	2.761	1.720	.208	7.651	3.479	2.958	1.118	2.372	.92
3:Archa	10.127	1.395	.000	6.726	3.354	1.227	8.886	8.118	4.227	.457	.835	2.44
4:Heineken	2.555	2.761	6.726	.000	2.760	2.817	6.555	.565	5.073	6.054	8.302	3.71
5:Kronenbourg	8.298	1.720	3.354	2.760	.000	1.268	9.504	4.984	4.048	2.432	6.177	4.86
6:Singha	5.553	.208	1.227	2.817	1.268	.000	6.195	3.704	4.074	.647	2.165	1.37
7:Budweiser light	7.977	7.651	8.886	6.555	9.504	6.195	.000	5.475	16.883	7.444	6.895	6.13
8:Budweiser	.903	3.479	8.118	.565	4.984	3.704	5.475	.000	7.228	7.238	8.475	3.14
9:Chang Classic	8.673	2.958	4.227	5.073	4.048	4.074	16.883	7.228	.000	5.675	7.378	4.96
10:Chang Export	9.370	1.118	.457	6.054	2.432	.647	7.444	7.238	5.675	.000	1.147	2.45
11:Chang Draught	9.532	2.372	.835	8.302	6.177	2.165	6.895	8.475	7.378	1.147	.000	1.85
12:LEO	3.110	.927	2.447	3.712	4.862	1.379	6.130	3.143	4.969	2.451	1.850	.00
13:Miller lite	10.598	5.998	4.414	10.828	10.941	5.261	3.786	9.691	14.160	4.312	1.660	3.71
14:Augsberger	4.494	11.406	20.127	6.522	13.221	13.423	22.440	5.849	10.581	19.505	22.266	11.70
15:Olympia Gold light	22.626	19.604	17.236	24.349	25.591	17.543	6.804	21.451	34.556	15.720	11.455	15.31
16:Heilemans old style	.628	6.917	13.450	4.917	11.606	8.479	12.857	2.800	9.975	12.872	12.815	4.95
17:Coors light	8.140	6.473	6.488	11.124	12.870	6.243	4.775	8.844	15.516	5.990	2.910	3.22
18:Hamms	1.100	3.304	7.245	4.053	7.734	3.829	5.946	2.159	9.580	6.156	5.707	1.48
19:Schiltz	2.257	1.791	5.345	4.306	5.689	2.694	9.805	3.349	6.093	4.552	5.038	1.22
20:Schiltz light	13.300	7.364	4.978	12.136	11.447	6.155	3.571	11.429	15.998	4.786	2.376	5.48

This is a dissimilarity matrix

ตาราง Proximity เป็นตารางแสดงความใกล้หรือระยะห่างของสมาชิกแต่ละคู่

Agglomeration Schedule

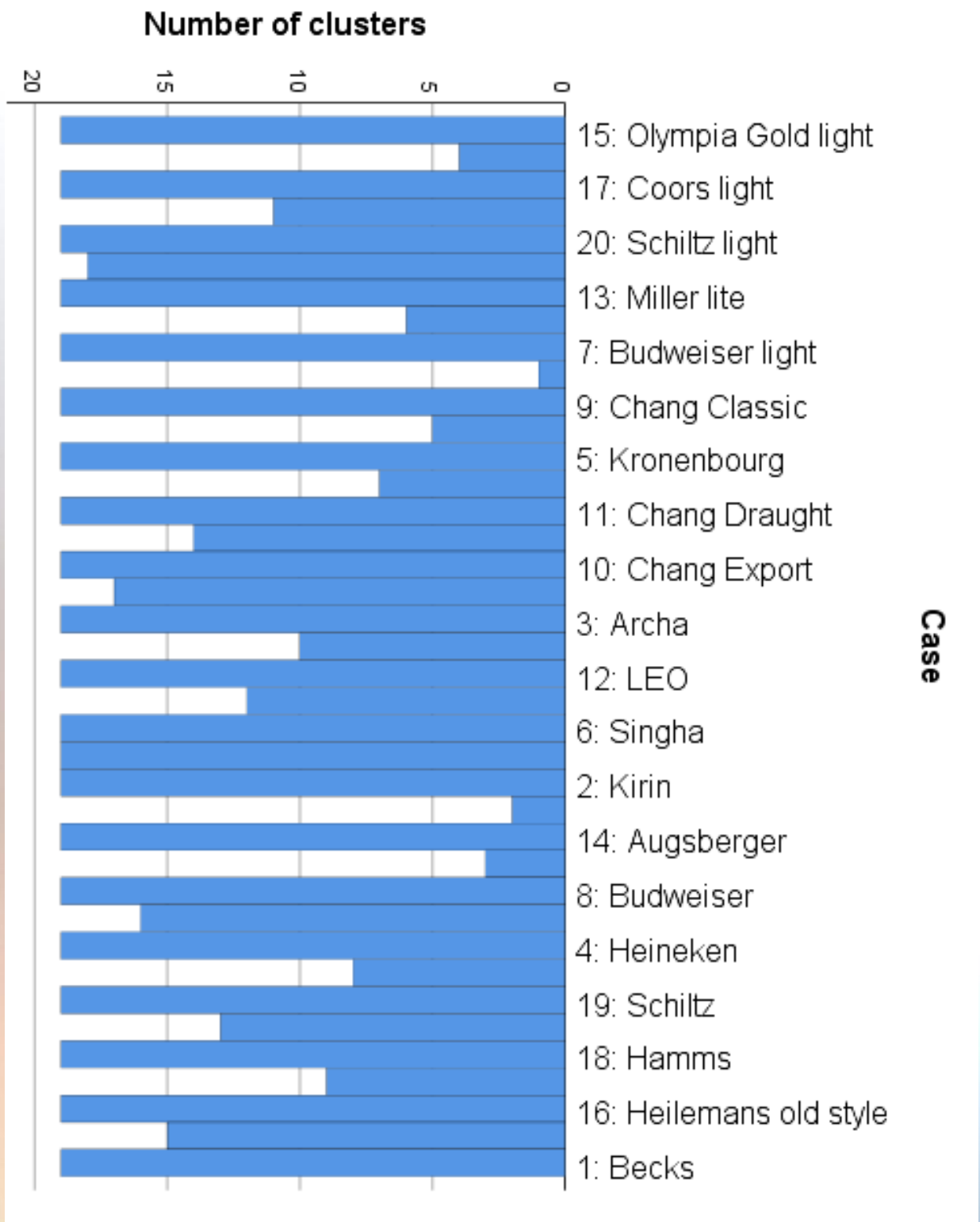
Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	2	6	.208	0	0	8
2	13	20	.322	0	0	9
3	3	10	.457	0	0	6
4	4	8	.565	0	0	12
5	1	16	.628	0	0	11
6	3	11	.991	3	0	10
7	18	19	1.037	0	0	11
8	2	12	1.153	1	0	10
9	13	17	1.540	2	0	14
10	2	3	1.741	8	6	13
11	1	18	2.122	5	7	12
12	1	4	3.130	11	4	17
13	2	5	3.302	10	0	15
14	7	13	4.044	0	9	16
15	2	9	4.761	13	0	18
16	7	15	5.228	14	0	19
17	1	14	6.113	12	0	18
18	1	2	7.583	17	15	19
19	1	7	12.169	18	16	0

ตาราง Agglomeration Schedule เป็นตารางแสดงการรวมกลุ่มแสดงให้เห็นถึงกระบวนการที่สมาชิก (Case) และกลุ่ม (Cluster) มารวมกันแต่ละขั้นตอน

Cluster Membership

Case	5 Clusters	4 Clusters	3 Clusters
1:Becks	1	1	1
2:Kirin	2	2	2
3:Archa	2	2	2
4:Heineken	1	1	1
5:Kronenbourg	2	2	2
6:Singha	2	2	2
7:Budweiser light	3	3	3
8:Budweiser	1	1	1
9:Chang Classic	2	2	2
10:Chang Export	2	2	2
11:Chang Draught	2	2	2
12:LEO	2	2	2
13:Miller lite	3	3	3
14:Augsberger	4	4	1
15:Olympia Gold light	5	3	3
16:Heilemans old style	1	1	1
17:Coors light	3	3	3
18:Hamms	1	1	1
19:Schiltz	1	1	1
20:Schiltz light	3	3	3

ตาราง Cluster Membership เป็นตารางที่ระบุว่าแต่ละสมาชิก (Case) จัดอยู่ในกลุ่ม (Cluster) ใดตามที่กำหนดไว้ในหน้าจอ Save New Variable ที่ผู้วิจัยกำหนดเอาไว้


Number of clusters
20
15
10
5
0
15: Olympia Gold light
17: Coors light
20: Schiltz light
13: Miller lite
7: Budweiser light
9: Chang Classic
5: Kronenbourg
11: Chang Draught
10: Chang Export
3: Archa
12: LEO
6: Singha
2: Kirin
14: Augsberger
8: Budweiser
4: Heineken
19: Schiltz
18: Hamms
16: Heilemans old style
1: Becks
Case

จากภาพ พื้นที่สีฟ้าแสดงการจัดกลุ่ม ถ้ามี 1 กลุ่ม (Cluster) พื้นที่จะเป็นสีฟ้าทั้งหมด ซึ่งจากตัวอย่างนี้จัดทั้งหมด 5 กลุ่ม พื้นที่สีฟ้าจะแสดง ดังนี้

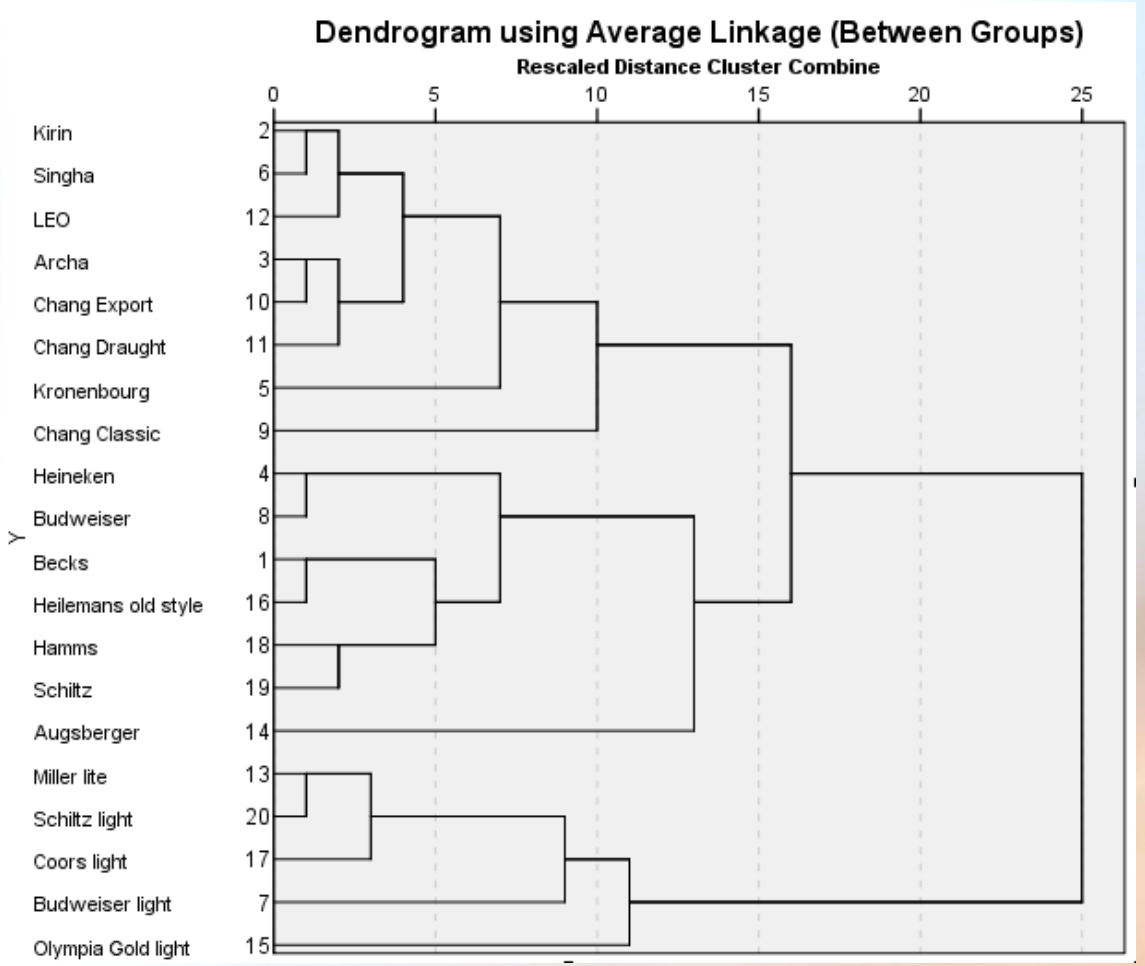
(1) Case ที่ 15 จะมีพื้นที่สีฟ้าเฉพาะตัว จึงมีสมาชิกตัวเดียวคือ Case ที่ 15

(2) Case ที่ 17 ถึง Case ที่ 7 ติดต่อจะมีพื้นที่สีฟ้า จึงเป็นสมาชิกกลุ่มเดียวกัน

(3) Case ที่ 9 ถึง Case ที่ 2 ติดต่อจะมีพื้นที่สีฟ้า จึงเป็นสมาชิกกลุ่มเดียวกัน

(4) Case ที่ 14 จะมีพื้นที่สีฟ้าเฉพาะตัว จึงมีสมาชิกตัวเดียวคือ Case ที่ 14

(5) Case ที่ 8 ถึง Case ที่ 1 ติดต่อจะมีพื้นที่สีฟ้า จึงเป็นสมาชิกกลุ่มเดียวกัน



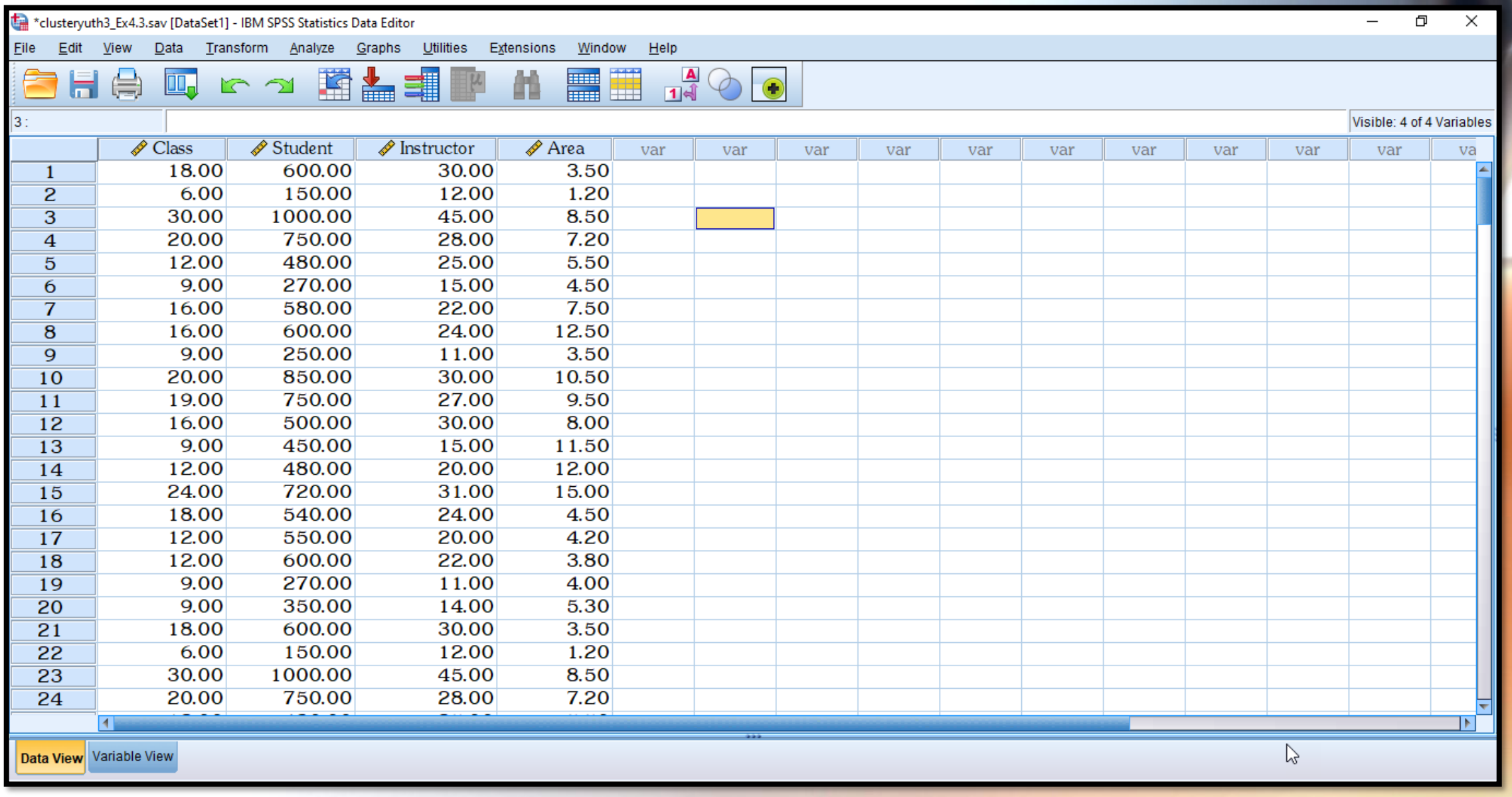
จากภาพ เป็นภาพ Dendrogram ซึ่งเป็นการแสดงการรวมสมาชิก (Case) ที่แสดงให้ดูอีกแบบหนึ่งที่เป็นการรวมสมาชิกออกมาโดยการโยงเส้นระหว่างสมาชิกที่รวมกลุ่มพร้อมทั้งบอกระยะห่างของการรวมกลุ่มด้วย โดยมีการแปลงระยะห่างให้อยู่ในช่วง 0 ถึง 25

ตัวอย่างเทคนิคการจัดกลุ่มแบบ K – Mean Cluster Analysis

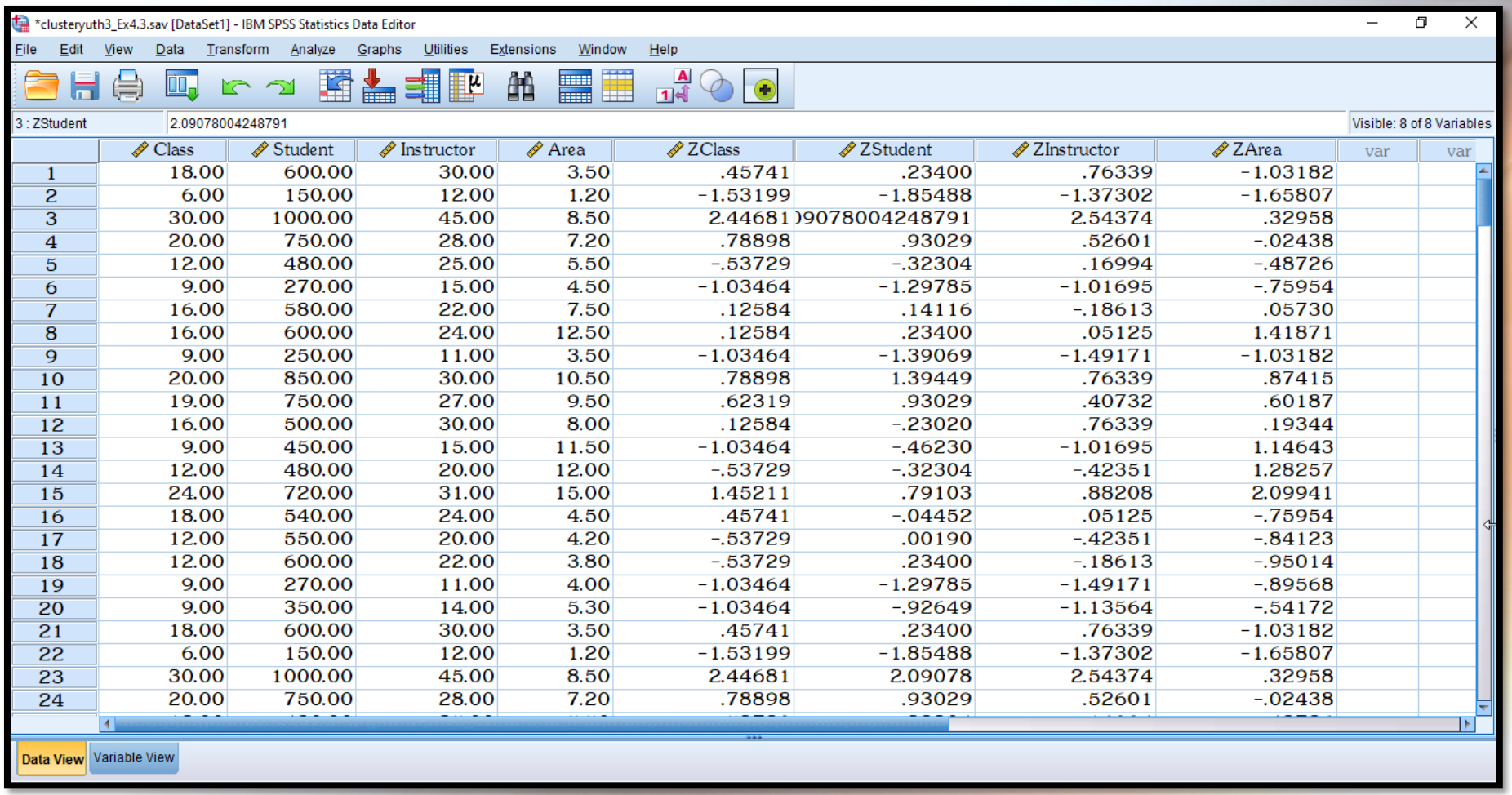
ตัวอย่าง จาก File clusteryuth1_Ex4.3 ผู้วิจัยจัดกลุ่มสมาชิก (Case) ที่เป็นโรงเรียนมัธยมศึกษาตอนต้น ในเขตพื้นที่การศึกษาซึ่งมีทั้งหมด 220 โรงเรียน เพื่อต้องการจัดกลุ่มออกเป็น 3 กลุ่ม จึงรวบรวมข้อมูลตัวแปรดังนี้

1. จำนวนห้องเรียน (Class)
2. จำนวนนักเรียน (Student)
3. จำนวนครู (Instructor)
4. พื้นที่ของโรงเรียน (จำนวนไร่) (Area)

ได้แฟ้มข้อมูลดังภาพประกอบ


*clusteryuth3_Ex4.3.sav [DataSet1] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help
3 :
Visible: 4 of 4 Variables
Class Student Instructor Area var var var var var var var var var var va
1 18.00 600.00 30.00 3.50
2 6.00 150.00 12.00 1.20
3 30.00 1000.00 45.00 8.50
4 20.00 750.00 28.00 7.20
5 12.00 480.00 25.00 5.50
6 9.00 270.00 15.00 4.50
7 16.00 580.00 22.00 7.50
8 16.00 600.00 24.00 12.50
9 9.00 250.00 11.00 3.50
10 20.00 850.00 30.00 10.50
11 19.00 750.00 27.00 9.50
12 16.00 500.00 30.00 8.00
13 9.00 450.00 15.00 11.50
14 12.00 480.00 20.00 12.00
15 24.00 720.00 31.00 15.00
16 18.00 540.00 24.00 4.50
17 12.00 550.00 20.00 4.20
18 12.00 600.00 22.00 3.80
19 9.00 270.00 11.00 4.00
20 9.00 350.00 14.00 5.30
21 18.00 600.00 30.00 3.50
22 6.00 150.00 12.00 1.20
23 30.00 1000.00 45.00 8.50
24 20.00 750.00 28.00 7.20
Data View
Variable View

ขั้นตอนที่ 1 ทำการ Standardized ทุกครั้งก่อนการวิเคราะห์การจัดกลุ่มแบบ K – Mean ด้วยคำสั่ง Analyze ➡ Descriptive Statistics ➡ Descriptive จากนั้นนำตัวแปรทั้ง 4 ตัวแปรใส่ในช่อง Variable(s) คลิก ☑ Save standardized value as variables จะได้ตัวแปรใหม่ดังภาพประกอบ

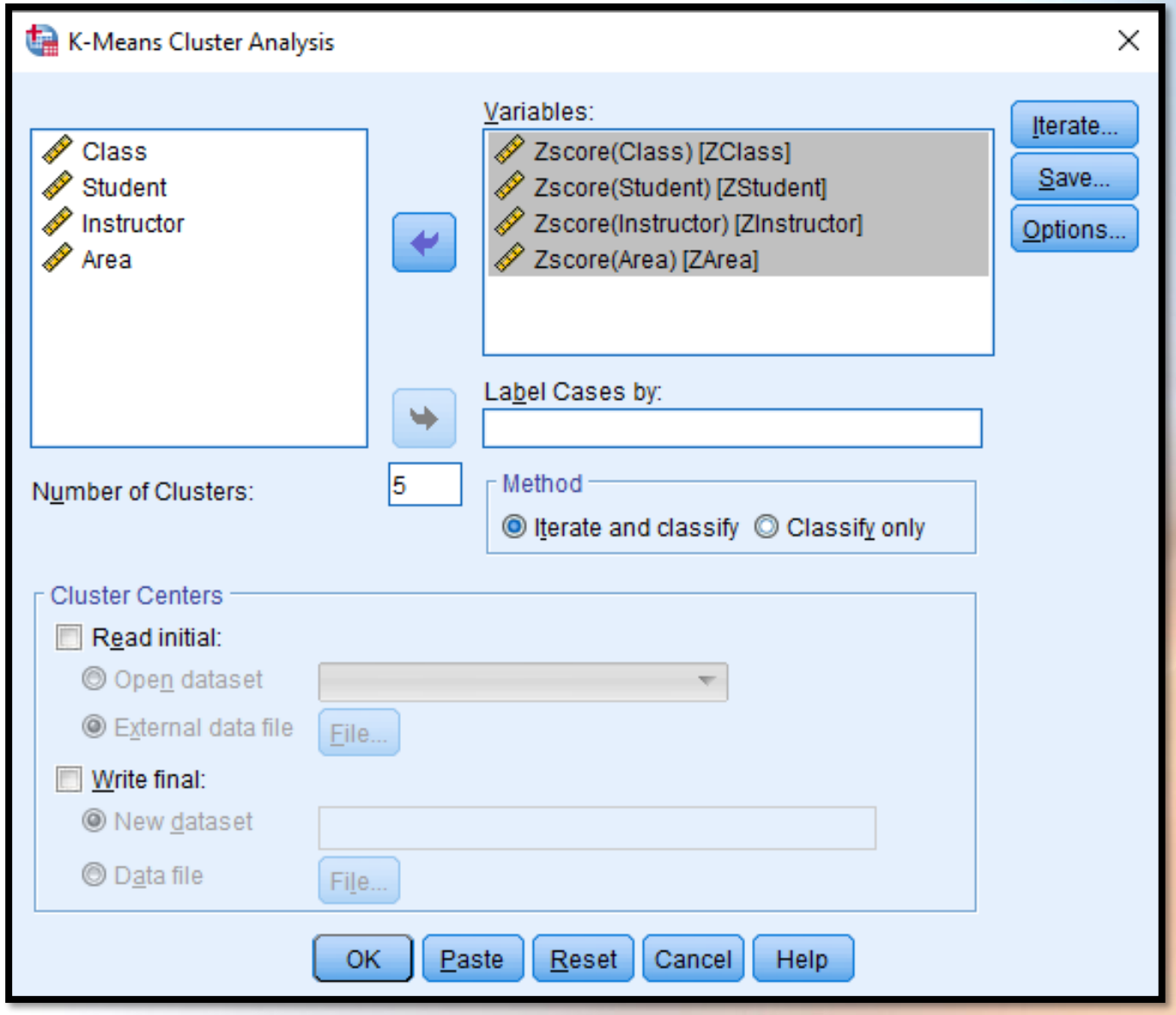


	Class	Student	Instructor	Area	ZClass	ZStudent	ZInstructor	ZArea	var	var
1	18.00	600.00	30.00	3.50	.45741	.23400	.76339	-1.03182		
2	6.00	150.00	12.00	1.20	-1.53199	-1.85488	-1.37302	-1.65807		
3	30.00	1000.00	45.00	8.50	2.44681	09078004248791	2.54374	.32958		
4	20.00	750.00	28.00	7.20	.78898	.93029	.52601	-.02438		
5	12.00	480.00	25.00	5.50	-.53729	-.32304	.16994	-.48726		
6	9.00	270.00	15.00	4.50	-1.03464	-1.29785	-1.01695	-.75954		
7	16.00	580.00	22.00	7.50	.12584	.14116	-.18613	.05730		
8	16.00	600.00	24.00	12.50	.12584	.23400	.05125	1.41871		
9	9.00	250.00	11.00	3.50	-1.03464	-1.39069	-1.49171	-1.03182		
10	20.00	850.00	30.00	10.50	.78898	1.39449	.76339	.87415		
11	19.00	750.00	27.00	9.50	.62319	.93029	.40732	.60187		
12	16.00	500.00	30.00	8.00	.12584	-.23020	.76339	.19344		
13	9.00	450.00	15.00	11.50	-1.03464	-.46230	-1.01695	1.14643		
14	12.00	480.00	20.00	12.00	-.53729	-.32304	-.42351	1.28257		
15	24.00	720.00	31.00	15.00	1.45211	.79103	.88208	2.09941		
16	18.00	540.00	24.00	4.50	.45741	-.04452	.05125	-.75954		
17	12.00	550.00	20.00	4.20	-.53729	.00190	-.42351	-.84123		
18	12.00	600.00	22.00	3.80	-.53729	.23400	-.18613	-.95014		
19	9.00	270.00	11.00	4.00	-1.03464	-1.29785	-1.49171	-.89568		
20	9.00	350.00	14.00	5.30	-1.03464	-.92649	-1.13564	-.54172		
21	18.00	600.00	30.00	3.50	.45741	.23400	.76339	-1.03182		
22	6.00	150.00	12.00	1.20	-1.53199	-1.85488	-1.37302	-1.65807		
23	30.00	1000.00	45.00	8.50	2.44681	2.09078	2.54374	.32958		
24	20.00	750.00	28.00	7.20	.78898	.93029	.52601	-.02438		

ขั้นตอนที่ 2 ทำการจัดกลุ่มแบบ K – Means ตามขั้นตอนดังนี้

1. Analyze ➡ Classify ➡ K – Means Cluster
2. นำตัวแปรที่ Standardized แล้วใส่ช่อง Variables
3. ที่ Label Cases By เป็นคำสั่งให้จัดกลุ่มตัวแปรในที่นี้คือ โรงเรียน (School) และตัวแปรนี้เป็นตัวแปรชนิด String และถ้าผู้วิเคราะห์ไม่กำหนด Label ของ Case โปรแกรมจะกำหนดลำดับ Case เป็นชื่อ Case ในที่นี้ไม่กำหนดตัวแปร
4. Number of Cluster พิมพ์เลขจำนวนกลุ่ม ในที่นี้จัด 5 กลุ่ม พิมพ์เลข 5
5. ที่ Method เลือก Iterate and Classify

ดังภาพประกอบ

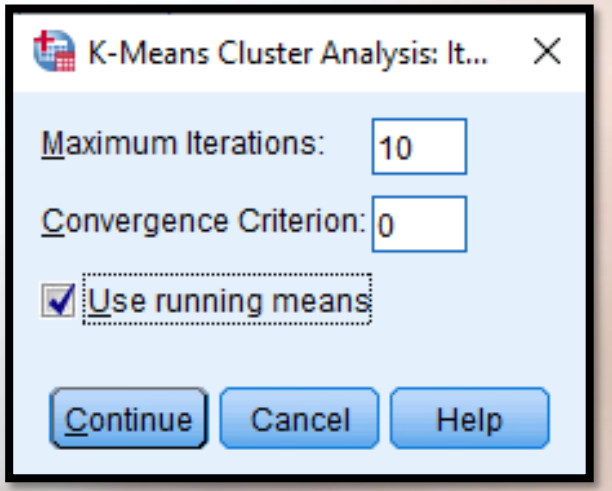

K-Means Cluster Analysis
Class
Student
Instructor
Area
Variables:
Zscore(Class) [ZClass]
Zscore(Student) [ZStudent]
Zscore(Instructor) [ZInstructor]
Zscore(Area) [ZArea]
Iterate...
Save...
Options...
Label Cases by:
Number of Clusters:
5
Method
Iterate and classify
Classify only
Cluster Centers
Read initial:
Open dataset
External data file
File...
Write final:
New dataset
Data file
File...
OK
Paste
Reset
Cancel
Help

ขั้นตอนที่ 3 ที่ Iterate กำหนด ดังนี้

3.1 Maximum Iteration ให้กำหนดจำนวนรอบในการคำนวณสูงสุด เลือกใส่ได้ 1 ถึง 999 ในที่นี้ใส่เลข 10 นั่นคือ ให้คำนวณ 10 รอบเท่านั้น โปรแกรมจะคำนวณไม่เกินจำนวนรอบที่กำหนด ในที่นี้กำหนด 10 รอบ

3.2 Convergence Criterion ใส่เลข 0 (ศูนย์) แต่ไม่เกิน (<) 1 คำสั่งนี้เป็นคำสั่งของระยะห่างที่สั้นที่สุดระหว่างค่าเฉลี่ยของกลุ่ม (Cluster) ในตอนเริ่มต้น

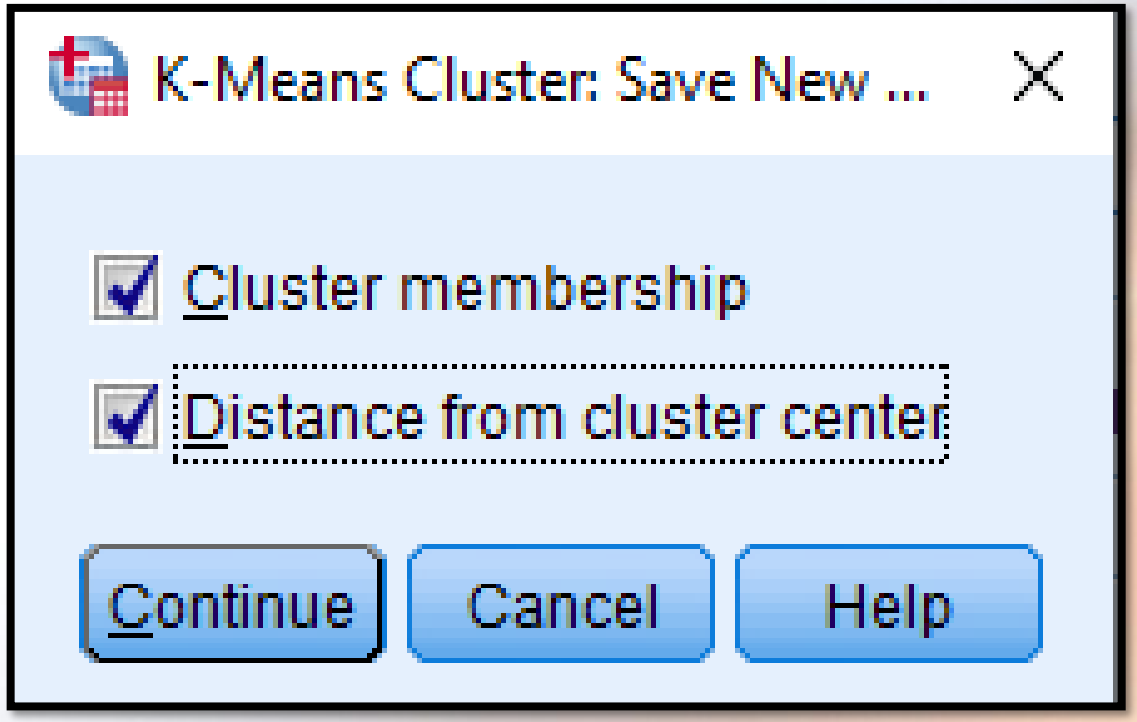
3.3 ☑ Use Running Means โปรแกรมจะแสดงค่าเฉลี่ยของกลุ่ม (Cluster) ทุกครั้งที่มีการกำหนดสมาชิกแก่กลุ่ม ดังภาพประกอบ



ขั้นตอนที่ 4 ที่ Save เลือก ดังนี้

4.1 เลือก ☑ Cluster Membership แสดงสมาชิกของกลุ่ม

4.2 เลือก ☑ Distance From Cluster Center แสดงค่าระยะห่าง Euclidean จากสมาชิกไปยังค่าเฉลี่ยของกลุ่ม ในที่นี้เลือกทั้ง 2 คำสั่ง ดังภาพประกอบ



ขั้นตอนที่ 5 ที่ Option เลือก ดังนี้

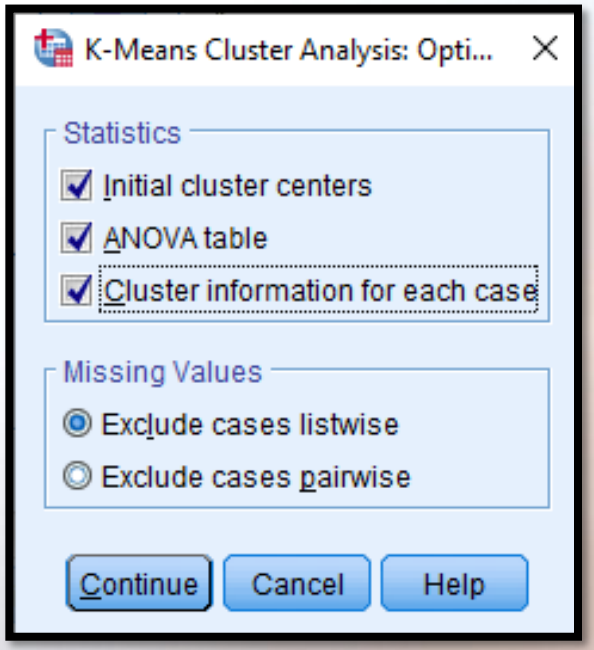
5.1 ที่ Statistics

5.1.1 ☑ Initial Cluster Centers

5.1.2 ☑ ANOVA Table

5.1.3 ☑ Cluster Information for Each Case

ดังภาพประกอบ



5.2 ที่ Missing Values มีคำสั่งย่อย 2 คำสั่งดังนี้

5.2.1 Exclude Case List wise เป็นคำสั่งให้ตัด Case ที่มี Missing value ออกจากการวิเคราะห์

5.2.2 Exclude Case Pair wise ที่มี Missing Values จับคู่กับ Case ใด จะถูกตัดออกจากการวิเคราะห์

ในที่นี้ให้เลือก Exclude Case List wise จากนั้นคลิก OK โปรแกรมจะแสดงผลลัพธ์ดังนี้

Initial Cluster Centers

	Cluster				
	1	2	3	4	5
Zscore(Class)	1.45211	-1.53199	2.44681	-1.03464	.45741
Zscore(Student)	.79103	-1.85488	2.09078	-.46230	.23400
Zscore(Instructor)	.88208	-1.37302	2.54374	-1.01695	.76339
Zscore(Area)	2.09941	-1.65807	.32958	1.14643	-1.03182

ตาราง Initial Cluster Centers เป็นตารางแสดงค่าเฉลรายของตัวแปรที่ Standardized แล้วและปรากฏใน Cluster ต่าง ๆ ทั้ง 5 Cluster ที่กำหนดไว้ตั้งแต่เริ่มต้น

Iteration History[a]

	Change in Cluster Centers				
Iteration	1	2	3	4	5
1	1.082	.802	4.441E-16	.820	.867
2	.030	.016	.000	.023	.010
3	.001	.000	.000	.001	.000
4	2.320E-5	6.817E-6	.000	1.757E-5	1.190E-6
5	6.444E-7	1.391E-7	.000	4.881E-7	1.322E-8
6	1.790E-8	2.839E-9	.000	1.356E-8	1.469E-10
7	4.972E-10	5.794E-11	.000	3.766E-10	1.632E-12
8	1.381E-11	1.183E-12	.000	1.046E-11	1.815E-14
9	3.838E-13	2.419E-14	.000	2.906E-13	2.498E-16
10	1.046E-14	5.439E-16	.000	8.049E-15	.000

a. Iterations stopped because the maximum number of iterations was performed. Iterations failed to converge. The maximum absolute coordinate change for any center is 8.660E-15. The current iteration is 10. The minimum distance between initial centers is 2.928.

ตาราง Iteration History แสดงค่าเฉลี่ยของ Cluster แต่ละรอบการคำนวณที่กำหนดเอาไว้ 10 รอบการคำนวณสูงสุด

Cluster Membership

Case Number	Cluster	Distance
1	5	.877
2	2	.819
3	3	4.441E-16
4	5	1.176
5	5	.779
6	2	.453
7	5	.656
8	4	.907
9	2	.245
10	1	.487
11	1	.711
12	5	.894
13	4	.843
14	4	.159
15	1	1.113
16	5	.550
17	5	1.009
18	5	.924
19	2	.322
20	2	.756
21	5	.877
22	2	.819
23	3	4.441E-16
24	5	1.176
25	5	.779
195	5	.894
196	4	.843
197	4	.159
198	1	1.113
199	5	.877
200	2	.819
201	3	4.441E-16
202	5	1.176
203	5	.779
204	2	.453
205	5	.656
206	4	.907
207	2	.245
208	1	.487
209	1	.711
210	5	.894
211	4	.843
212	4	.159
213	1	1.113
214	5	.550
215	5	.877
216	2	.819
217	3	4.441E-16
218	5	1.176
219	5	.779
220	2	.453

ตาราง Cluster Membership จะแสดงทุก Case ที่นำมาวิเคราะห์โดยจะแสดงให้ทราบว่าแต่ละ Cluster นั้นมี Case ใด เช่น Case ที่ 1 อยู่ใน Cluster ที่ 5 และมีระยะห่างจากค่าเฉลี่ยของ Cluster ที่ 5 = 0.877

Final Cluster Centers

	Cluster				
	1	2	3	4	5
Zscore(Class)	.94055	-1.16934	2.44681	-.48045	.09232
Zscore(Student)	1.04568	-1.43324	2.09078	-.17980	.12708
Zscore(Instructor)	.67861	-1.29390	2.54374	-.46420	.24196
Zscore(Area)	1.16588	-1.06245	.32958	1.28257	-.43800

ตาราง Final Cluster Centers เป็นค่าเฉลี่ยของตัวแปรที่ Standardized แล้ว ค่าเฉลี่ยนี้ก็คือค่าเฉลี่ยของแต่ละ Cluster จากตารางพบว่า ตัวแปร Class จะมีค่าเฉลี่ยแตกต่างกันในแต่ละ Cluster โดยที่ตัวแปร Class ใน Cluster ที่ 1 = .94055 และใน Cluster ที่ 5 = .09232 นั่นคือ Class ใน Cluster ที่ 1 มีค่ามากกว่าค่าเฉลี่ยรวมเท่ากับ .94055 และ Class ใน Cluster ที่ 5 มีค่ามากกว่าค่าเฉลี่ยรวมเท่ากับ .09232

Distances between Final Cluster Centers

Cluster	1	2	3	4	5
1		4.411	2.746	2.200	2.080
2	4.411		6.493	2.869	2.603
3	2.746	6.493		4.866	3.910
4	2.200	2.869	4.866		1.970
5	2.080	2.603	3.910	1.970	

ตาราง Distance between Final Cluster Centers เป็นตารางแสดงระยะห่างระหว่างค่าเฉลี่ยของ Cluster ทั้ง 5 Clusters จากตารางจะเห็นว่า Cluster ที่ 3 ห่างจาก Cluster ที่ 2 = 6.493 ซึ่งถือว่าเป็นระยะห่างที่มากที่สุด และระหว่าง Cluster ที่ 4 กับ Cluster ที่ 4 กับ Cluster ที่ 5 มีระยะห่างสั้นที่สุดซึ่งมีระยะห่าง = 1.970

ANOVA

	Cluster		Error			
	Mean Square	df	Mean Square	df	F	Sig.
Zscore(Class)	45.816	4	.166	215	275.630	.000
Zscore(Student)	49.067	4	.106	215	464.059	.000
Zscore(Instructor)	48.337	4	.119	215	405.134	.000
Zscore(Area)	44.455	4	.192	215	232.086	.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.

ตาราง ANOVA เป็นการทดสอบความแตกต่างค่าเฉลี่ยของแต่ละตัวแปร เมื่อถูกจัดอยู่ต่าง Cluster กัน จากตารางพบว่า ทุกตัวแปรเมื่ออยู่ต่าง Cluster กัน มีความแตกต่างกัน โดยดูจากค่า Sig = .000 เมื่อพิจารณารายตัวแปร พบว่า ตัวแปร Student เมื่ออยู่ต่าง Cluster กันจะมีความแตกต่างกันมากที่สุด (F = 464.059) ส่วนตัวแปร Area เมื่ออยู่ต่าง Cluster กันจะมีความแตกต่างกันน้อยที่สุด (F = 232.086)

Number of Cases in each Cluster

Cluster	1	35.000
	2	48.000
	3	13.000
	4	35.000
	5	89.000
Valid		220.000
Missing		.000

ตาราง Number of Case in each Cluster แสดงจำนวนข้อมูล Cases ที่นำมาจัดกลุ่มทั้ง 220 Cases โดยจัดอยู่ Cluster ต่าง ๆ ดังนี้

Cluster ที่ 1 = 35 Cases

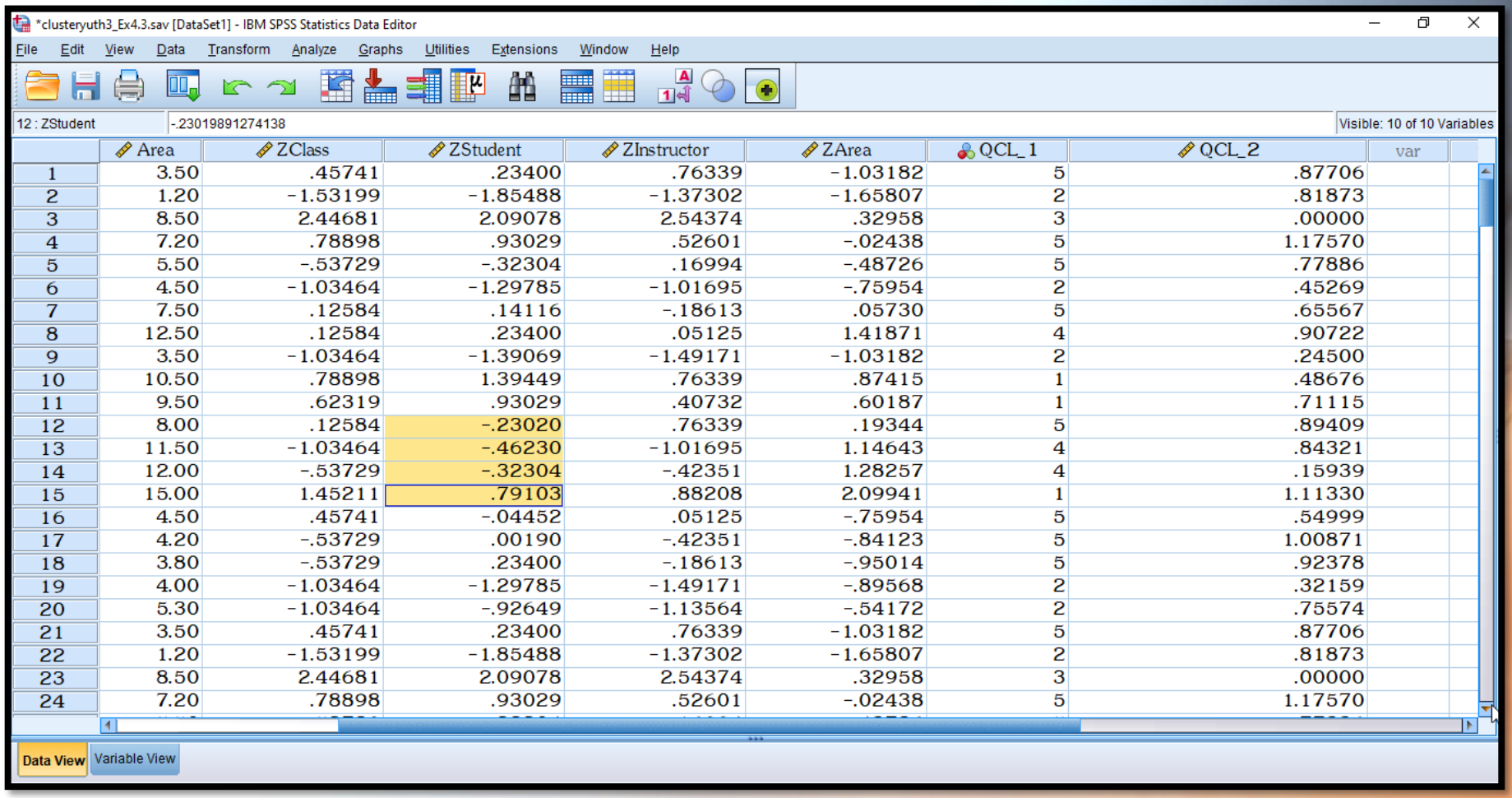
Cluster ที่ 2 = 48 Cases

Cluster ที่ 3 = 13 Cases

Cluster ที่ 4 = 35 Cases

Cluster ที่ 5 = 89 Cases

รวม 220 Cases และไม่มี Missing (Missing = .000)



	Area	ZClass	ZStudent	ZInstructor	ZArea	QCL_1	QCL_2	var
1	3.50	.45741	.23400	.76339	-1.03182	5	.87706	
2	1.20	-1.53199	-1.85488	-1.37302	-1.65807	2	.81873	
3	8.50	2.44681	2.09078	2.54374	.32958	3	.00000	
4	7.20	.78898	.93029	.52601	-.02438	5	1.17570	
5	5.50	-.53729	-.32304	.16994	-.48726	5	.77886	
6	4.50	-1.03464	-1.29785	-1.01695	-.75954	2	.45269	
7	7.50	.12584	.14116	-.18613	.05730	5	.65567	
8	12.50	.12584	.23400	.05125	1.41871	4	.90722	
9	3.50	-1.03464	-1.39069	-1.49171	-1.03182	2	.24500	
10	10.50	.78898	1.39449	.76339	.87415	1	.48676	
11	9.50	.62319	.93029	.40732	.60187	1	.71115	
12	8.00	.12584	-.23020	.76339	.19344	5	.89409	
13	11.50	-1.03464	-.46230	-1.01695	1.14643	4	.84321	
14	12.00	-.53729	-.32304	-.42351	1.28257	4	.15939	
15	15.00	1.45211	.79103	.88208	2.09941	1	1.11330	
16	4.50	.45741	-.04452	.05125	-.75954	5	.54999	
17	4.20	-.53729	.00190	-.42351	-.84123	5	1.00871	
18	3.80	-.53729	.23400	-.18613	-.95014	5	.92378	
19	4.00	-1.03464	-1.29785	-1.49171	-.89568	2	.32159	
20	5.30	-1.03464	-.92649	-1.13564	-.54172	2	.75574	
21	3.50	.45741	.23400	.76339	-1.03182	5	.87706	
22	1.20	-1.53199	-1.85488	-1.37302	-1.65807	2	.81873	
23	8.50	2.44681	2.09078	2.54374	.32958	3	.00000	
24	7.20	.78898	.93029	.52601	-.02438	5	1.17570	

จาก Data View โปรแกรมจะแสดง Case ว่าอยู่ใน Cluster ใด โดยดูจาก Column QCL_1 เช่น Case ที่ 1 อยู่ใน Cluster ที่ 5 ส่วน QCL_2 เป็นการแสดงระยะห่างแต่ละ Case ไปยังค่าเฉลี่ยของ Cluster นั่นคือ Case ที่ 1 มีระยะห่างจากค่าเฉลี่ยของ Cluster = 0.87706