# แผนบริหารการสอนประจำบทที่ 3

วิชา 4114601 การวิเคราะห์สถิติขั้นสูงด้วยโปรแกรมสำเร็จรูปทางสถิติ
แผนบริหารการสอน บทที่ 3 การจำแนกกลุ่มตัวแปรด้วยเทคนิค Cluster Analysis
เวลา 8 ชั่วโมง

**สาระสำคัญ**

Cluster Analysis (CA) เป็นเทคนิคที่ใช้จำแนกหรือแบ่งหน่วยสังเกตหรือหน่วยที่ให้ข้อมูล (Case) ออกเป็นกลุ่มย่อย ๆ ตั้งแต่ 2 กลุ่มขึ้นไป โดยหลักการของ CA ก็คือ การจัดกลุ่มให้กับ Case ที่อยู่ในกลุ่มเดียวกันมีลักษณะที่เหมือนกันหรือคล้ายคลึงกันให้ได้มากที่สุด ส่วน Case ที่อยู่ต่างกลุ่มกันมีลักษณะที่แตกต่างกันให้ได้มากที่สุด ดังนั้น การพิจารณาเลือกลักษณะหรือตัวแปรที่จะนำมาใช้ในการแบ่งกลุ่ม Case จึงมีความสำคัญ ซึ่งการจำแนกกลุ่มตัวแปรด้วยเทคนิค Cluster Analysis เป็นเทคนิคที่ใช้ในการจำแนกกลุ่มโดยไม่ทราบมาก่อนว่าควรมีกี่กลุ่ม แต่จะแบ่งตามค่าของตัวแปรที่นำมาใช้ในการแบ่ง โดยให้หน่วยที่อยู่ในกลุ่มเดียวกัน มีความคล้ายกันในตัวแปรที่ศึกษา แต่หน่วยที่อยู่ต่างกลุ่มกันจะมีความต่างกัน

**ผลการเรียนรู้ที่คาดหวัง**

นักศึกษาสามารถเข้าใจหลักการจำแนกกลุ่มตัวแปรด้วยเทคนิค Cluster Analysis และเห็นข้อแตกต่างระหว่างเทคนิค Factor Analysis กับ Cluster Analysis รวมทั้งสามาถประยุกต์ใช้เทคนิค Cluster Analysis ในงานด้านต่างๆได้

**จุดประสงค์การเรียนรู้**

1. เข้าใจหลักการจำแนกกลุ่มตัวแปรด้วยเทคนิค Cluster Analysis
2. เปรียบเทียบความแตกต่างระหว่างเทคนิค Factor Analysis กับ Cluster Analysis ได้
3. สามาถประยุกต์ใช้เทคนิค Cluster Analysis ในงานด้านต่างๆได้
4. สามารถสรุปผลที่ได้จากการวิเคราะห์ด้วยเทคนิค Cluster Analysis ได้อย่างถูกต้อง

**กิจกรรมการเรียนการสอน**

1. นำเสนอ PowerPoint เนื้อหาเกี่ยวกับจำแนกกลุ่มตัวแปรด้วยเทคนิค Cluster Analysis จากเอกสารประกอบการสอนการวิเคราะห์สถิติขั้นสูงด้วยโปรแกรมสำเร็จรูปทางสถิติ
2. ให้นักศึกษาค้นคว้าข้อมูลงานวิจัยที่เลือกใช้การจำแนกกลุ่มตัวแปรด้วยเทคนิค Cluster Analysis ในงานวิจัย และนำมาอภิปรายในชั้นเรียน
3. ให้นักศึกษาทำแบบฝึกหัดท้ายบท

**สื่อการเรียนรู้**

1. ตัวอย่างงานวิจัย
2. PowerPoint
3. เอกสารประกอบการสอนรายวิชา 4114601 การวิเคราะห์สถิติขั้นสูงด้วยโปรแกรมสำเร็จรูปทางสถิติ

**การวัดและประเมินผล**

1. ประเมินผลจากแบบฝึกหัด
2. ประเมินผลจากการถาม – ตอบในชั้นเรียน
3. ประเมินผลจากการทดสอบย่อยรายบท

# บทที่ 3
# การจำแนกกลุ่มตัวแปรด้วยเทคนิค Cluster Analysis

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

## 1. วัตถุประสงค์ของ Cluster Analysis

เทคนิค Cluster Analysis มีวัตถุประสงค์หลักที่สำคัญ คือ การจัดกลุ่มหน่วยวิเคราะห์ ซึ่งจะเป็นประโยชน์ในงานด้านต่าง ๆ เช่น ด้านการตลาด ด้านการแพทย์ ด้านการปกครอง เป็นต้น ดังตัวอย่างต่อไปนี้

**ตัวอย่างที่ 1** ใช้ศึกษาพฤติกรรมการบริโภคของกลุ่มผู้บริโภคที่อยู่ต่างกลุ่มกัน ซึ่งจะทำให้สามารถวางกลยุทธ์ทางการตลาดได้อย่างมีประสิทธิภาพมากขึ้น การที่จะสามารถแยกกลุ่มผู้บริโภคออกเป็นกลุ่มย่อยได้ จะต้องพิจารณาถึงตัวแปรที่ใช้ในการแบ่งกลุ่มผู้บริโภค ที่จะทำให้ผู้ที่อยู่ต่างกลุ่มกันมีพฤติกรรมการบริโภคที่แตกต่างกัน ตัวแปรดังกล่าวอาจจะประกอบด้วยอาชีพ อายุ รายได้ เป็นต้น

**ตัวอย่างที่ 2** ใช้แบ่งกลุ่มประเทศเป็นกลุ่มประเทศพัฒนาแล้ว กลุ่มประเทสกำลังพัฒนา และกลุ่มประเทศด้อยพัฒนา โดยอาจจะพิจารณาจากตัวแปรลักษณะทางเศรษฐกิจ ดัชนีการพัฒนามนุษย์ การส่งเสริมคุณภาพชีวิตและสาธารณสุข รายได้ประชากร เป็นต้น

**ตัวอย่างที่ 3** การเปรียบเทียบรถยนต์ยี่ห้อต่าง ๆ โดยที่ 1 Case คือ รถยนต์ 1 ยี่ห้อ ซึ่งพิจารณาจากตัวแปร เช่น ความถี่ในการซ่อม ลูกสูบ ระบบเบรก ค่าใช้จ่ายต่อกิโลเมตร ราคา เป็นต้น

## 2. ประเภทของเทคนิค Cluster Analysis

เทคนิค Cluster Analysis แบ่งเป็นหลายประเภทหรือเทคนิคย่อย โดยเทคนิคที่ใช้กันมากมี 2 เทคนิค คือ เทคนิค Hierarchical Cluster Analysis และ K-Means Cluster Analysis

2.1 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

สำหรับขั้นตอนของเทคนิค Hierarchical Cluster ในการแบ่งกลุ่ม Case มีดังนี้คือ

<u>ขั้นที่ 1</u> เลือกตัวแปรหรือปัจจัยที่คาดว่ามีอิทธิพลที่ทำให้ Case ต่างกัน นั่นคือ ตัวแปร

นั้นจะทำให้สามารถแบ่งกลุ่ม Case ได้ชัดเจน ขั้นตอนนี้เป็นขั้นตอนที่สำคัญ

ขั้นที่ 2 เลือกวิธีการวัดระยะห่างระหว่าง Case แต่ละคู่ หรือเลือกวิธีการคำนวณเพื่อวัดค่าความคล้ายของ Case แต่ละคู่

ขั้นที่ 3 เลือกหลักเกณฑ์ในการรวมกลุ่ม หรือรวม Cluster

2.2 เทคนิค K-Means Cluster Analysis

เป็นเทคนิคการจำแนก Case ออกเป็นกลุ่มย่อย จะใช้เมื่อมีจำนวน Case มาก โดยจะต้องกำหนดจำนวนกลุ่ม หรือจำนวน Cluster 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 ไม่มากรควรใช้เทคนิค Hierarchical

2. เทคนิค K-Means นั้น ผู้ใช้จะต้องกำหนดจำนวนกลุ่มที่แน่นอนไว้ล่วงหน้า กรณีที่ผู้วิเคราะห์ยังไม่แน่ใจว่าควรมีกี่กลุ่มจึงจะเหมาะสม ผู้วิเคราะห์อาจจะใช้วิธีใดวิธีหนึ่งดังต่อไปนี้

❖ ทำการวิเคราะห์ด้วยวิธี K-Means หลาย ๆ ครั้ง แต่ละครั้งกำหนดจำนวนกลุ่มแตกต่างกันไป เช่น เป็น 3, 4 หรือ 5 กลุ่ม แล้วพิจารณาหาจำนวนกลุ่มที่เหมาะสม แต่เมื่อมีข้อมูลมากวิธีนี้จะทำให้เสียเวลามาก

❖ ใช้ข้อมูลบางส่วนทำการวิเคราะห์โดยวิธี Hierarchical เพื่อหาจำนวนกลุ่มที่ควรจะเป็นจากนั้นจึงใช้เทคนิค K-Means กับข้อมูลทั้งหมดที่มี

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

4. วิธี K-Means จะหาระยะห่างโดยวิธี Euclidean distance โดยอัตโนมัติ ขณะที่ Hierarchical ผู้วิเคราะห์มีสิทธิ์ที่จะเลือกวิธีการคำนวณระยะห่าง หรือความคล้ายได้

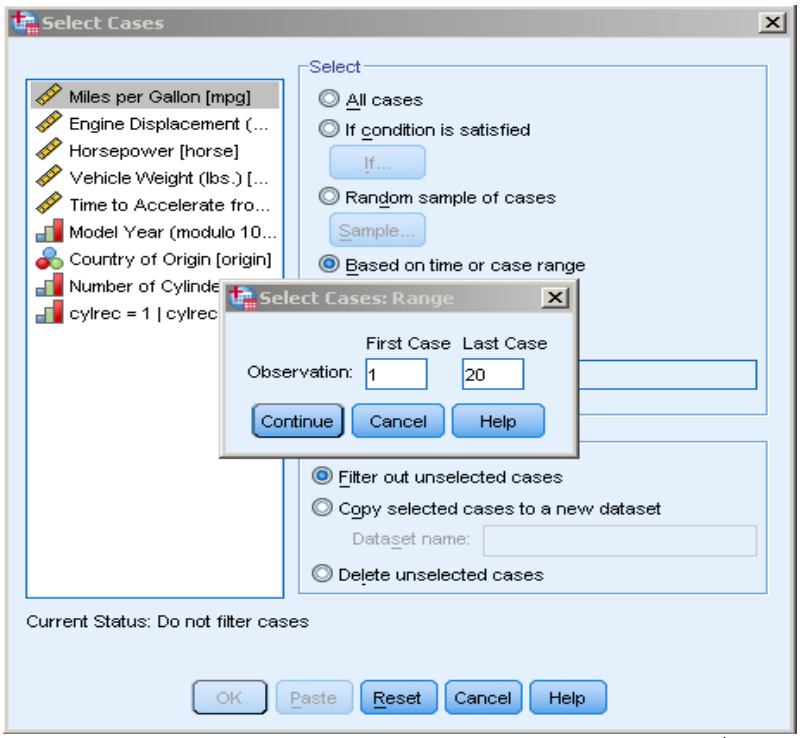
**3. ตัวอย่างการจำแนกกลุ่มตัวแปรด้วยเทคนิค Cluster Analysis โดยใช้โปรแกรมโปรแกรมสำเร็จรูปทางสถิติ**

**ตัวอย่างที่ 1** ตัวอย่างการใช้เทคนิค Hierarchical Cluster Analysis

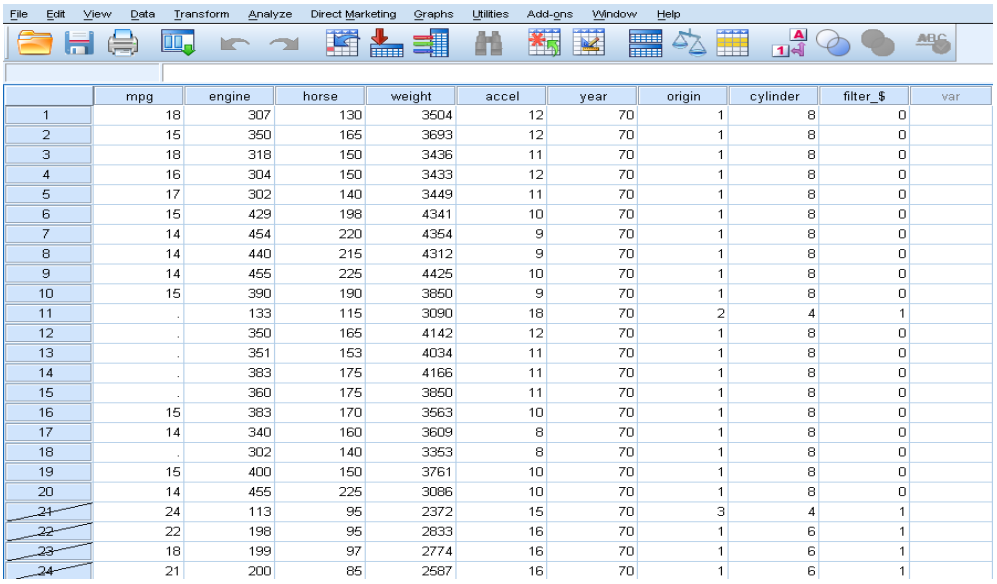
สำหรับตัวอย่างนี้จะใช้ข้อมูลแค่ 20 Caseแรกในการจัดกลุ่ม เนื่องจากไม่ต้องการให้ผลลัพธ์ที่ได้ยาวเกินไป จนทำให้ไม่สะดวกในการอธิบายความหมาย โดยมีขั้นตอนดังนี้

1. เปิดไฟล์ Cars.sav ในโปรแกรมสำเร็จรูปทางสถิติ
2. เลือก Case ที่ 1 – 20 เพื่อใช้ในการวิเคราะห์ โดยใช้คำสั่ง

Data ⟶ Select Cases ... Click Based on time or case range จากนั้น Click range ใส่เลข 1 ที่ First Case และเลข 20 ที่ Last Case จะได้หน้าจอดังภาพ



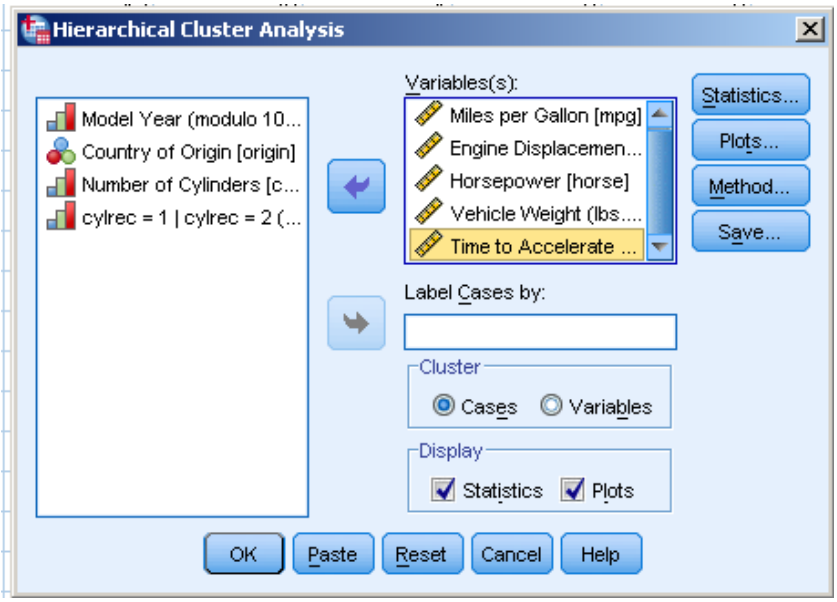

3. Click Continue แล้ว OK โปรแกรมจะเลือกเฉพาะ Case ที่ 1 ถึง Case ที่ 20 มาวิเคราะห์ดังภาพ



| | mpg | engine | horse | weight | accel | year | origin | cylinder | filter_$ | var |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 18 | 307 | 130 | 3504 | 12 | 70 | 1 | 8 | 0 | |
| 2 | 15 | 350 | 165 | 3693 | 12 | 70 | 1 | 8 | 0 | |
| 3 | 18 | 318 | 150 | 3436 | 11 | 70 | 1 | 8 | 0 | |
| 4 | 16 | 304 | 150 | 3433 | 12 | 70 | 1 | 8 | 0 | |
| 5 | 17 | 302 | 140 | 3449 | 11 | 70 | 1 | 8 | 0 | |
| 6 | 15 | 429 | 198 | 4341 | 10 | 70 | 1 | 8 | 0 | |
| 7 | 14 | 454 | 220 | 4354 | 9 | 70 | 1 | 8 | 0 | |
| 8 | 14 | 440 | 215 | 4312 | 9 | 70 | 1 | 8 | 0 | |
| 9 | 14 | 455 | 225 | 4425 | 10 | 70 | 1 | 8 | 0 | |
| 10 | 15 | 390 | 190 | 3850 | 9 | 70 | 1 | 8 | 0 | |
| 11 | . | 133 | 115 | 3090 | 18 | 70 | 2 | 4 | 1 | |
| 12 | . | 350 | 165 | 4142 | 12 | 70 | 1 | 8 | 0 | |
| 13 | . | 351 | 153 | 4034 | 11 | 70 | 1 | 8 | 0 | |
| 14 | . | 383 | 175 | 4166 | 11 | 70 | 1 | 8 | 0 | |
| 15 | . | 360 | 175 | 3850 | 11 | 70 | 1 | 8 | 0 | |
| 16 | 15 | 383 | 170 | 3563 | 10 | 70 | 1 | 8 | 0 | |
| 17 | 14 | 340 | 160 | 3609 | 8 | 70 | 1 | 8 | 0 | |
| 18 | . | 302 | 140 | 3353 | 8 | 70 | 1 | 8 | 0 | |
| 19 | 15 | 400 | 150 | 3761 | 10 | 70 | 1 | 8 | 0 | |
| 20 | 14 | 455 | 225 | 3086 | 10 | 70 | 1 | 8 | 0 | |
| 21 | 24 | 113 | 95 | 2372 | 15 | 70 | 3 | 4 | 1 | |
| 22 | 22 | 198 | 95 | 2833 | 16 | 70 | 1 | 6 | 1 | |
| 23 | 18 | 199 | 97 | 2774 | 16 | 70 | 1 | 6 | 1 | |
| 24 | 21 | 200 | 85 | 2587 | 16 | 70 | 1 | 6 | 1 | |

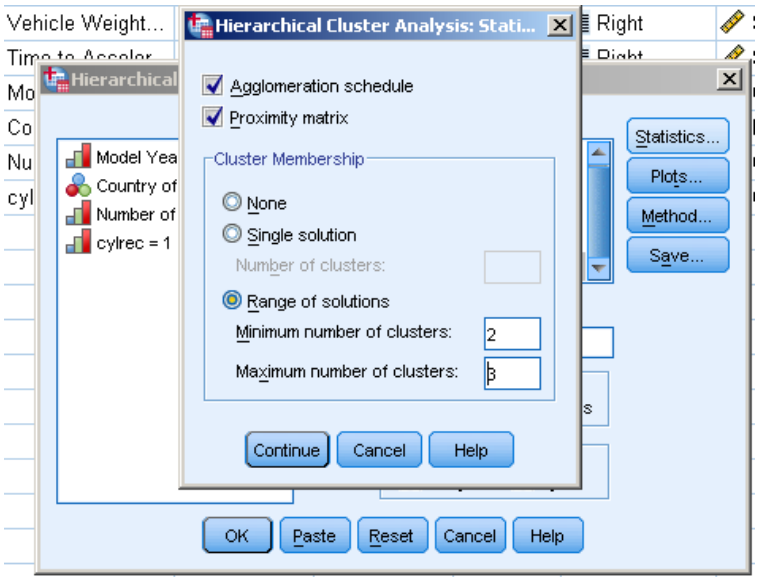
4. ทำการแบ่งกลุ่มด้วยเทคนิค Hierarchical Cluster โดยใช้คำสั่ง

Analyze⟶Classify ⟶ Hierarchical Cluster เลือกตัวแปร mpg ,engine ,horse ,weight และ accel ไว้ที่ช่อง Variables(s) : ดังภาพ

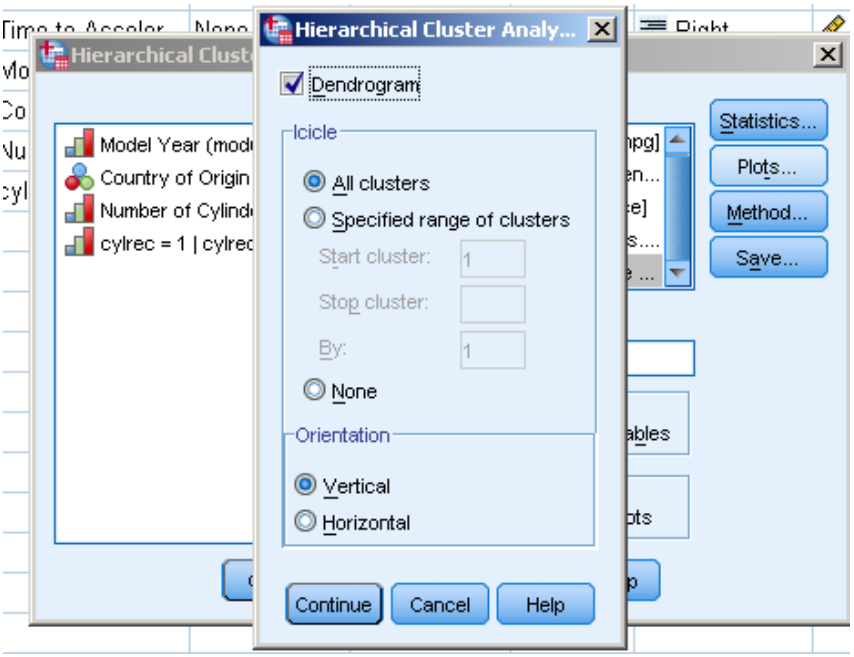



5. ในส่วนของ Cluster เลือก Cases เนื่องจากต้องการจัดกลุ่ม (Case) และในส่วนของ Display เลือก Statistics และ Plot

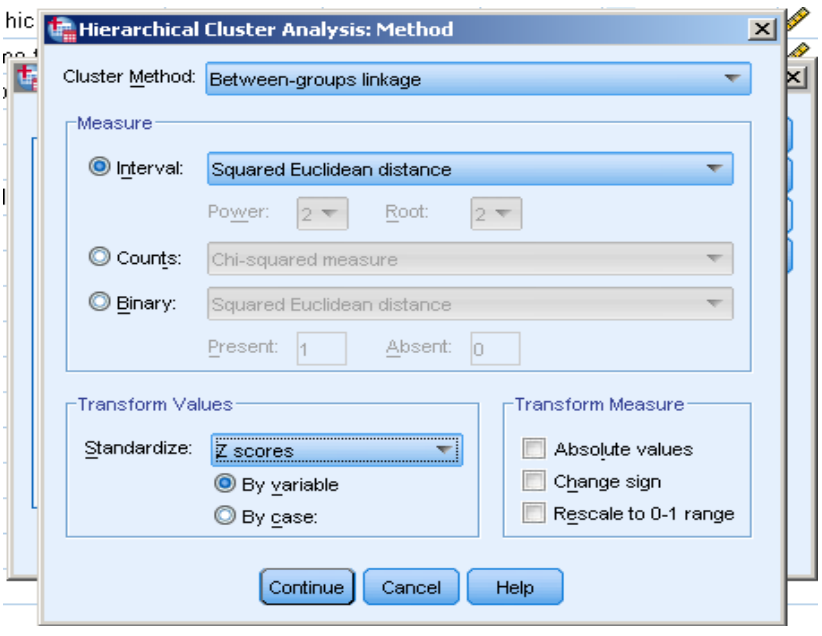
6. Click Statistics เลือก Agglomeration schedule และ Proximity matrix และ Click Range of solution ในส่วนของ Minimum number of clusters ใส่ 2 และ Maximum number of clusters ใส่ 3 ดังภาพ



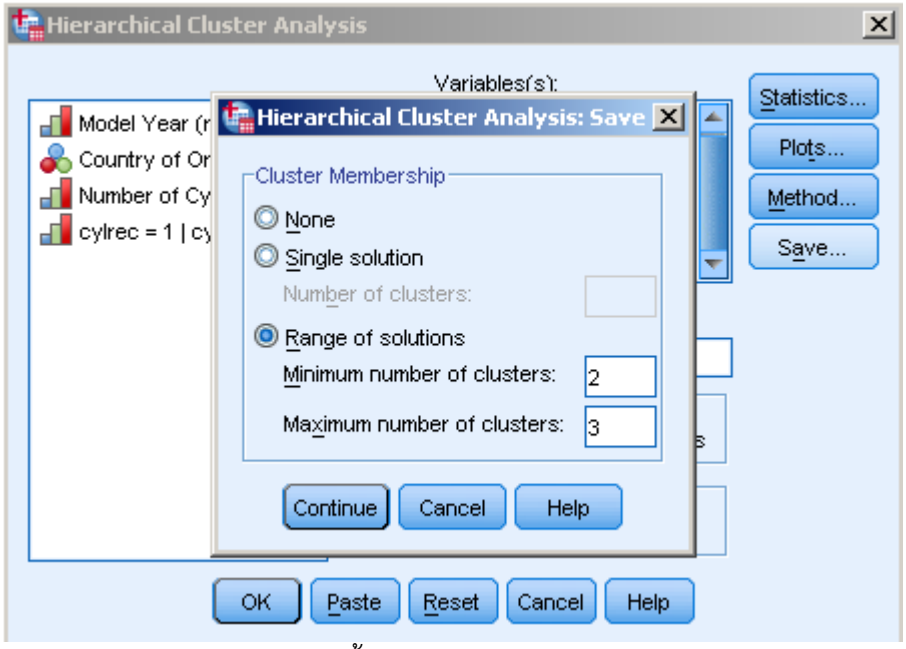
7. Click Continue จากนั้น Click Plot จากนั้น Click Dendrogram และ All Cluster แล้ว Continue ดังภาพ




8. Click Method ในส่วน Cluster Method เลือก Between – groups Linkage ในส่วนของ Measure เลือก Interval เนื่องจากตัวแปรทั้ง 5 ตัวที่เลือก เป็นข้อมูล Ratio scale และเลือก Square Euclidean distance และในส่วนของ Transform Values เลือก Z scores เนื่องจากตัวแปรทั้ง 4 ตัว ข้างต้นมีหน่วยที่แตกต่างกัน และเลือก By Variable จากนั้น Continue ดังภาพ



9. Click Save เลือก Range of solution Minimum number of clusters ใส่ 2 และ Maximum number of clusters ใส่ 3 จากนั้น Continue ดังภาพ




จะได้ผลลัพธ์จากการวิเคราะห์ดังนี้

**ตารางที่ 3.1** ตารางแสดงจำนวน Case ที่นำมาวิเคราะห์ และ จำนวน Missing

**Case Processing Summary**[a]

| Cases | | | | | |
|---|---|---|---|---|---|
| Valid | | Missing | | Total | |
| N | Percent | N | Percent | N | Percent |
| 14 | 70.0% | 6 | 30.0% | 20 | 100.0% |

a. Squared Euclidean Distance used

จากตารางพบว่า จากข้อมูล 20 Case มีค่า Missing อยู่ 6 จึงมีจำนวนCaseนำมา วิเคราะห์เพียง 14 หรือคิดเป็น 70% (14/20)

**ตารางที่ 3.2** ตารางแสดงระยะห่างของ Case แต่ละคู่

**Proximity Matrix**

| Case | Squared Euclidean Distance | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1:Case 1 | 2:Case 2 | 3:Case 3 | 4:Case 4 | 5:Case 5 | 6:Case 6 | 7:Case 7 | 8:Case 8 | 9:Case 9 | 10:Case 10 | 11:Case 16 | 12:Case 17 | 13:Case 19 | 14:Case 20 |
| 1:Case 1 | .000 | 6.302 | 1.024 | 2.319 | 1.974 | 18.914 | 30.418 | 30.160 | 28.953 | 17.610 | 9.841 | 18.698 | 11.307 | 25.208 |
| 2:Case 2 | 6.302 | .000 | 5.360 | 1.800 | 4.071 | 6.407 | 12.413 | 12.580 | 11.079 | 6.643 | 1.796 | 8.074 | 3.368 | 10.148 |
| 3:Case 3 | 1.024 | 5.360 | .000 | 2.603 | .797 | 14.979 | 24.340 | 23.700 | 23.971 | 11.987 | 6.585 | 13.644 | 8.191 | 19.213 |
| 4:Case 4 | 2.319 | 1.800 | 2.603 | .000 | 1.952 | 13.892 | 22.737 | 22.598 | 21.153 | 12.417 | 5.115 | 12.353 | 7.471 | 16.350 |
| 5:Case 5 | 1.974 | 4.071 | .797 | 1.952 | .000 | 13.962 | 22.351 | 21.198 | 22.681 | 9.635 | 4.779 | 9.079 | 5.848 | 18.117 |
| 6:Case 6 | 18.914 | 6.407 | 14.979 | 13.892 | 13.962 | .000 | 1.698 | 2.155 | 1.354 | 3.197 | 4.645 | 9.387 | 4.288 | 10.073 |
| 7:Case 7 | 30.418 | 12.413 | 24.340 | 22.737 | 22.351 | 1.698 | .000 | .239 | .663 | 3.981 | 8.180 | 10.484 | 7.722 | 9.576 |
| 8:Case 8 | 30.160 | 12.580 | 23.700 | 22.598 | 21.198 | 2.155 | .239 | .000 | 1.599 | 2.917 | 7.669 | 8.347 | 6.953 | 9.887 |
| 9:Case 9 | 28.953 | 11.079 | 23.971 | 21.153 | 22.681 | 1.354 | .663 | 1.599 | .000 | 5.954 | 8.723 | 13.552 | 8.887 | 9.970 |
| 10:Case 10 | 17.610 | 6.643 | 11.987 | 12.417 | 9.635 | 3.197 | 3.981 | 2.917 | 5.954 | .000 | 2.202 | 2.448 | 2.094 | 7.362 |
| 11:Case 16 | 9.841 | 1.796 | 6.585 | 5.115 | 4.779 | 4.645 | 8.180 | 7.669 | 8.723 | 2.202 | .000 | 3.547 | .804 | 5.856 |
| 12:Case 17 | 18.698 | 8.074 | 13.644 | 12.353 | 9.079 | 9.387 | 10.484 | 8.347 | 13.552 | 2.448 | 3.547 | .000 | 3.079 | 11.370 |
| 13:Case 19 | 11.307 | 3.368 | 8.191 | 7.471 | 5.848 | 4.288 | 7.722 | 6.953 | 8.887 | 2.094 | .804 | 3.079 | .000 | 8.969 |
| 14:Case 20 | 25.208 | 10.148 | 19.213 | 16.350 | 18.117 | 10.073 | 9.576 | 9.887 | 9.970 | 7.362 | 5.856 | 11.370 | 8.969 | .000 |

This is a dissimilarity matrix

จากตาราง พบว่า ค่าต่าง ๆ ในตาราง เป็นระยะห่างของ Case แต่ละคู่โดยระยะห่างที่ใช้คือ ค่า Squared Euclidean Distance เช่น case 1 และ case 9 ห่างกัน 28.963 ขณะที่ case 1 และ case 3 ห่างกันเพียง 1.004 ดังนั้น ควรจัด case 1 และ case 3 ให้อยู่ในกลุ่มเดียวกัน นั่นคือ case 1 และ case 3 มีค่าตัวแปร 5 ตัว ดังกล่าวคล้ายกัน ในขณะเดียวกัน ควรจัด case 1 และ case 9 อยู่ต่างกลุ่มกัน หรือ case 1 และ case 9 มีความแตกต่างกันในตัวแปรทั้ง 5 ตัว

**ตารางที่ 3.3** ตารางแสดงผลจากการใช้วิธี Between – groups linkage (หน้าจอ Method) ในการรวมกลุ่ม Case

**Average Linkage (Between Groups)**

**Agglomeration Schedule**

| Stage | Cluster Combined | | Coefficients | Stage Cluster First Appears | | Next Stage |
|---|---|---|---|---|---|---|
| | Cluster 1 | Cluster 2 | | Cluster 1 | Cluster 2 | |
| 1 | 7 | 8 | .239 | 0 | 0 | 4 |
| 2 | 3 | 5 | .797 | 0 | 0 | 5 |
| 3 | 11 | 13 | .804 | 0 | 0 | 8 |
| 4 | 7 | 9 | 1.131 | 1 | 0 | 6 |
| 5 | 1 | 3 | 1.499 | 0 | 2 | 10 |
| 6 | 6 | 7 | 1.735 | 0 | 4 | 11 |
| 7 | 2 | 4 | 1.800 | 0 | 0 | 10 |
| 8 | 10 | 11 | 2.148 | 0 | 3 | 9 |
| 9 | 10 | 12 | 3.025 | 8 | 0 | 11 |
| 10 | 1 | 2 | 3.768 | 5 | 7 | 13 |
| 11 | 6 | 10 | 7.180 | 6 | 9 | 12 |
| 12 | 6 | 14 | 9.133 | 11 | 0 | 13 |
| 13 | 1 | 6 | 14.933 | 10 | 12 | 0 |

จากผลลัพธ์ในตาราง เป็นผลจากการใช้วิธี Between – groups linkage (หน้าจอ Method) ในการรวมกลุ่ม case นั่นคือ ในแต่ละ stage จะบอกว่ามีการรวม case คู่ใดบ้าง ให้อยู่ในกลุ่มเดียวกัน เช่น
Stage 1 : จะจัดที่ 7 และ case ที่ 8 อยู่ในกลุ่มเดียวกัน เนื่องจาก caseที่ 7 และ 8 มีระยะห่างกันสั้นที่สุด ซึ่งระยะห่าง (ค่า Squared Euclidean Distance) คือค่าใน column ของ coefficients ซึ่งเท่ากับ .239 และค่า Next stage ใน column สุดท้าย = 4 หมายถึง กลุ่มหรือ cluster ที่มี caseที่ 7 และ 8 จะรวมกับ case อื่นต่อไปใน stage ที่ 4
Stage 2 : มีการจัดให้ caseที่ 3 และ caseที่ 5 ให้อยู่ในกลุ่มหรือ cluster เดียวกัน ซึ่ง caseที่ 3 และ 5 มีระยะห่าง = .797 และกลุ่มที่มี caseที่ 3 และ 5 อยู่จะรวมกับ case อื่นอีกใน stage ที่ 5 (Next stage = 5) เป็นต้น

ดังที่ได้กล่าวแล้วว่าเทคนิค Cluster ในขั้นแรกจะให้ จำนวนกลุ่ม = จำนวน Case นั่น คือ ในตัวอย่างนี้มี 14 Case (เนื่องจากมีการ Missing 6 Case) จึงเริ่มต้นมี 14 กลุ่ม ๆ ละ 1 Case แล้วจึงค่อย ๆ รวม Case ทีละคู่ ดังในตารางที่ จนในที่สุดเหลือกลุ่มเดียว ดังนั้น การ พิจารณาว่าควรแบ่งเป็นกี่กลุ่มย่อยจึงอยู่ที่การพิจารณาของผู้วิเคราะห์ โดยจะพิจารณาจากระยะห่าง หรือความคล้าย

**ตารางที่ 3.4** ตารางแสดงการระบุว่าแต่ละ Case อยู่กลุ่มใด

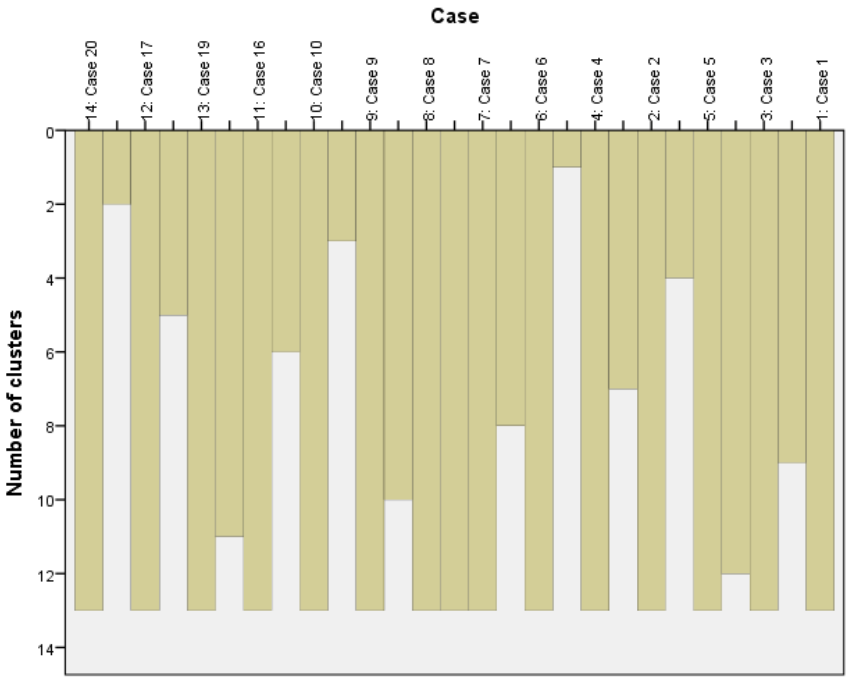
**Cluster Membership**

| | 3 Clusters | 2 Clusters |
|---|---|---|
| 1:Case 1 | 1 | 1 |
| 2:Case 2 | 1 | 1 |
| 3:Case 3 | 1 | 1 |
| 4:Case 4 | 1 | 1 |
| 5:Case 5 | 1 | 1 |
| 6:Case 6 | 2 | 2 |
| 7:Case 7 | 2 | 2 |
| 8:Case 8 | 2 | 2 |
| 9:Case 9 | 2 | 2 |
| 10:Case 10 | 2 | 2 |
| 11:Case 16 | 2 | 2 |
| 12:Case 17 | 2 | 2 |
| 13:Case 19 | 2 | 2 |
| 14:Case 20 | 3 | 2 |

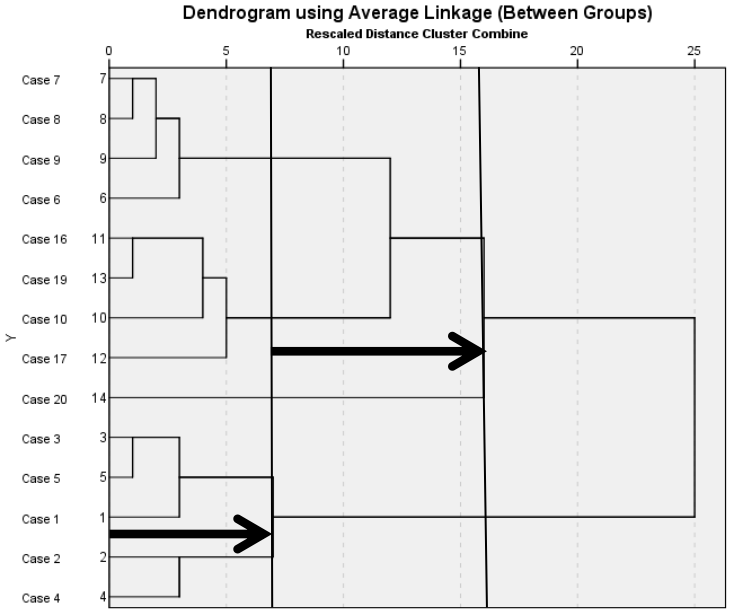
จากตารางเป็นการระบุว่าแต่ละ Case อยู่กลุ่มใดโดยแบ่งเป็น 2 รูปแบบ ดังนี้

1. กรณีที่มี 3 กลุ่ม คือ
   กลุ่มที่ 1 ประกอบไปด้วย case ที่ 1,2,3,4,5
   กลุ่มที่ 2 ประกอบไปด้วย case ที่ 6,7,8,9,10,16,17,19
   กลุ่มที่ 3 ประกอบไปด้วย case ที่ 20

2. กรณีที่แบ่งเป็น 2 กลุ่ม คือ
   กลุ่มที่ 1 ประกอบไปด้วย case ที่ 1,2,3,4,5
   กลุ่มที่ 2 ประกอบไปด้วย case ที่ 6,7,8,9,10,16,17,19,20




**รูปที่ 3.1** รูปแสดงการรวม Case โดยใช้วิธี Hierarchical Cluster Analysis



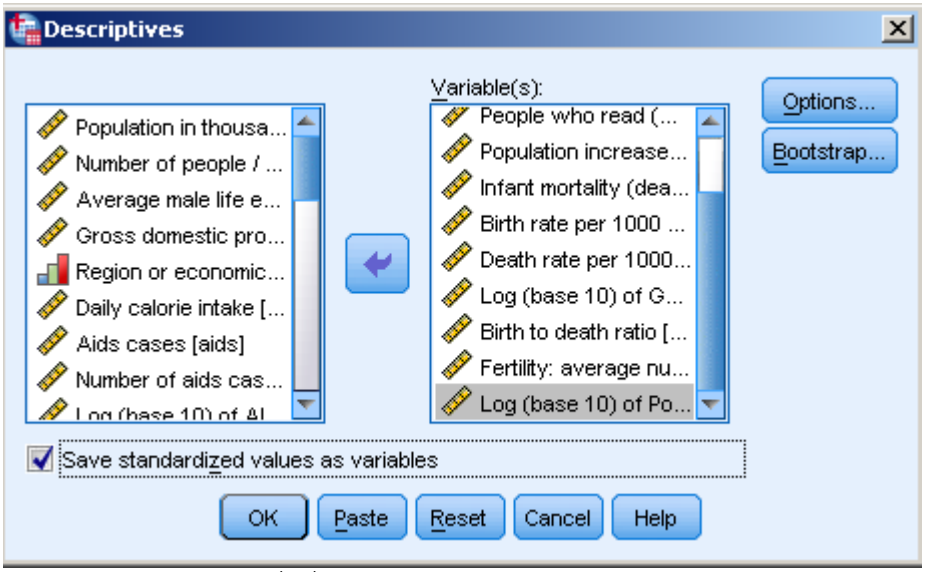

**รูปที่ 3.2** รูป Dendrogram แสดงการรวม Case โดยใช้วิธี Hierarchical Cluster Analysis

**ตัวอย่างที่ 2** ตัวอย่างการใช้เทคนิค K-Means Clustering

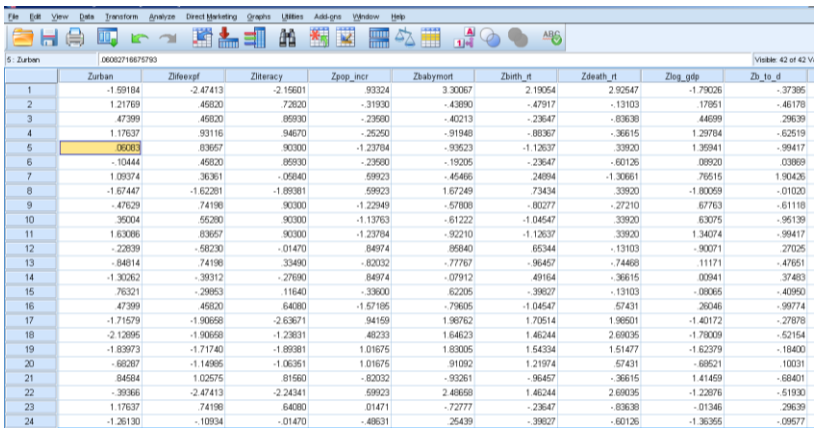
ในตัวอย่างนี้จะใช้แฟ้มข้อมูลที่มีอยู่ในโปรแกรมสำเร็จรูปทางสถิติ คือแฟ้มข้อมูล World 95 for Missing Values ถึงแม้แฟ้ม World 95 for Missing Values จะมีจำนวน case น้อยกว่า 200 แต่ก็มากพอที่จะใช้วิธี K-Means ได้ แฟ้ม World 95 for Missing Values เป็นแฟ้มแสดงตัวแปรต่าง ๆ ของแต่ละประเทศจำนวน 109 ประเทศ มีขั้นตอนการวิเคราะห์ดังนี้

1. เปิดไฟล์ World 95 for Missing Values .sav ในโปรแกรมสำเร็จรูปทางสถิติ
2. ทำการ Standardized ตัวแปรที่นำมาวิเคราะห์ ดังนี้

Analyze ⟶ Descriptive Statistics ⟶ Descriptives เลือกตัวแปร 11 ตัว คือ urban, lifeexpf, literacy, pop_incr, babymort, birth_rt, death_rt, log_gdp, b_to_d, fertility และ log_pop ใส่ใน box ของ Variable (s) (ตัวแปรทั้ง 11 ตัว เป็นตัวแปรชนิดตัวเลข) Click Save standardized value as variables จะได้หน้าจอดังภาพ

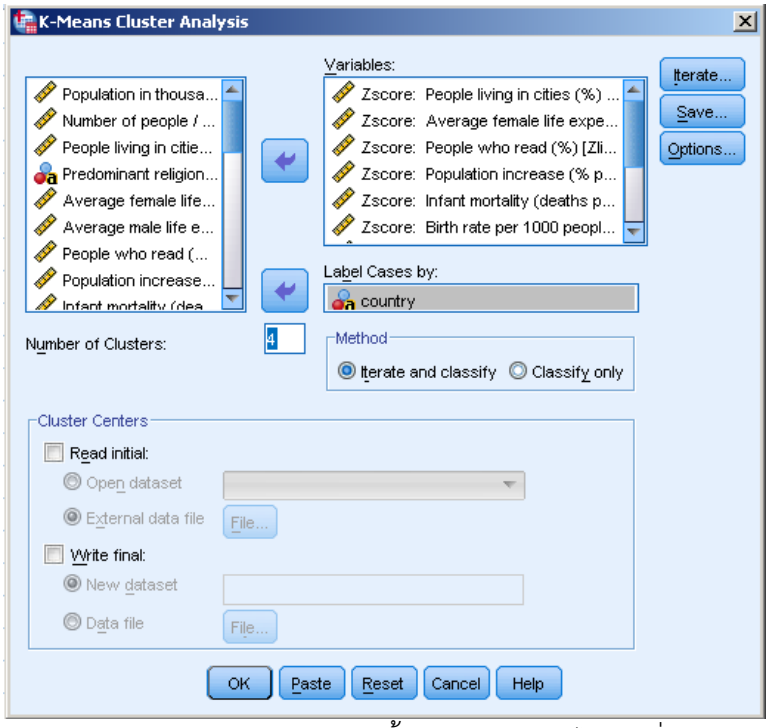



จะได้ตัวแปรใหม่ 11 ตัวที่มีชื่อเดิมแต่มี Z นำหน้าต่อจากตัวแปรสุดท้ายในแฟ้มข้อมูล ดังภาพ



3. จำแนกกลุ่มด้วยเทคนิค K-Means โดยใช้คำสั่ง

Analyze ⟶ Classify ⟶ K – Mean Cluster เลือกตัวแปร Zurban, Zlifeexpf, Zliteracy, Zpop_incr, Zbabymort, Zbirth_rt, Zdeath_rt, Zlog_gdp, Zb_to_d, Zfertility และ Zlog_pop ไว้ที่ช่อง Variables(s) : และเลือกตัวแปร country ใส่ไว้ในช่อง Label Case by : ส่วนในช่อง Number of Cluster ให้ใส่ 4 และ Method เลือก Iterate and classify ดังภาพ



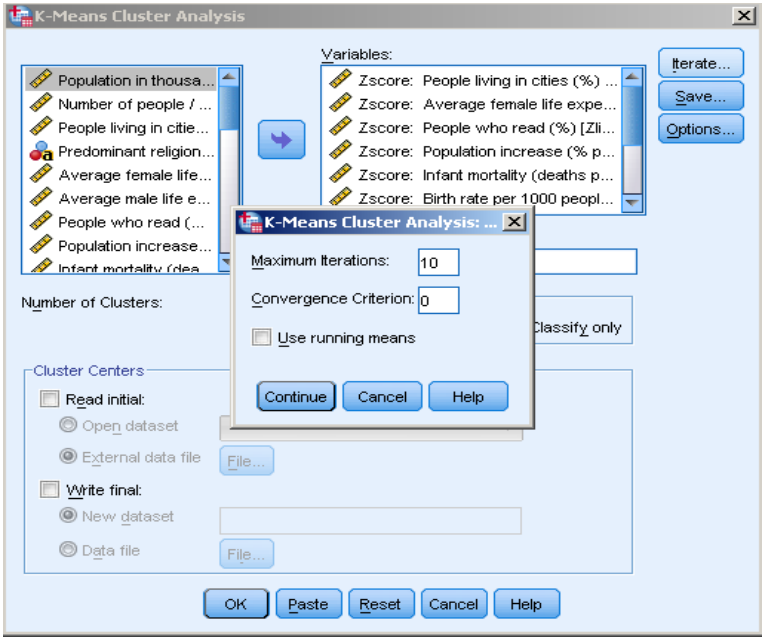

4. Click Iterate โดยในส่วนนี้จะสามารถเลือกได้ต่อเมื่อผู้วิเคราะห์เลือกวิธี Iterate and classify โดยจะมีให้เลือกอยู่ 3 ส่วน คือ

4.1 ในส่วนของ Maximum Iteration เป็นการกำหนดจำนวนรอบ (Iteration) ที่ใช้ในการคำนวณ ซึ่งตัวเลขที่ใส่ใน box ต้องมีค่าตั้งแต่ 1 ถึง 999 โดยโปรแกรมจะคำนวณไม่เกินจำนวนรอบที่กำหนด ในที่นี้ให้ใส่เลข 10

4.2 ในส่วนของ Convergence Criterion เป็นการกำหนดการหยุดการคำนวณ โดยการกำหนดสัดส่วนของระยะห่างที่สั้นที่สุด ระหว่างค่ากลางของ Cluster ในตอนเริ่มแรก โดยค่าที่กำหนดใน box จะต้องมากกว่า 0 แต่ไม่เกิน 1 ในที่นี้ให้ใส่เลข 0

4.3 ในส่วนของ Use running means ถ้าเลือกทางเลือกนี้หมายถึง จะให้หาค่ากลางของ Cluster ทุกครั้งที่มีการกำหนด Case ให้แก่ Cluster ถ้าไม่เลือกจะมีการคำนวณค่ากลางใหม่ต่อเมื่อได้กำหนด Cluster ให้แก่ทุก Case แล้ว ในที่นี้ไม่ต้องเลือก

จะได้หน้าจอดังภาพ



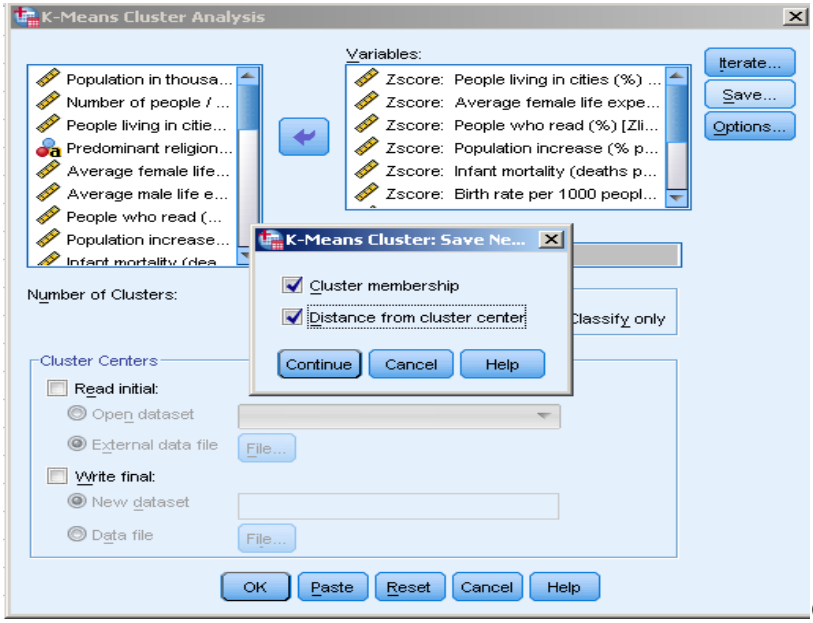

Click Continue

5. Click Save ในส่วนนี้ จะมีให้เลือกอยู่ 2 ส่วน คือ

5.1 Cluster membership จะสร้างค่าตัวแปรใหม่ซึ่งเป็นตัวแปรที่ระบุกลุ่มคือ Cluster ที่แต่ละ case เป็นสมาชิกอยู่

5.2 Distance from cluster center จะสร้างตัวแปรใหม่ โดยตัวแปรใหม่นี้จะระบุค่า Euclidean distance จากแต่ละ case ไปยังค่ากลางของกลุ่ม

โดยในตัวอย่างนี้ให้เลือกทั้ง 2 ส่วนคือ Cluster membership และ Distance from cluster center ดังภาพ




Click Continue

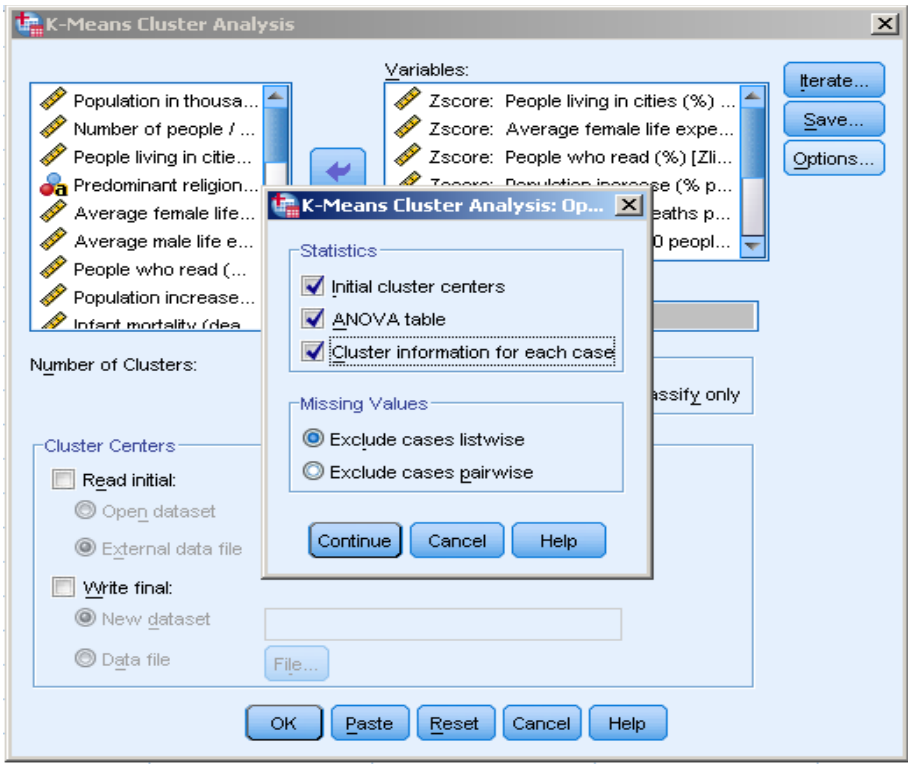
6. Click Option ในส่วนนี้ จะมีให้เลือกอยู่ 2 ส่วน คือ

6.1 Statistics มี 3 ทางเลือกคือ

- ❖ Initial cluster centers เป็นการให้แสดงค่ากลางของแต่ละกลุ่มในตอนเริ่มแรก
- ❖ ANOVA Table ให้ค่าสถิติ F เพื่อแสดงความแตกต่างระหว่างกลุ่มของตัวแปรแต่ละตัวเมื่ออยู่ต่างกลุ่มกัน
- ❖ Cluster information for each case จะแสดงรายละเอียดของ Cluster ให้สำหรับแต่ละ Case โดยในตัวอย่างนี้ให้เลือกทั้ง 3 ทางเลือก

6.2 Missing Values มีทางเลือกสำหรับค่า Missing คือ

- ❖ Exclude cases listwise
- ❖ Exclude cases pairwise โดยในตัวอย่างนี้ให้เลือก Exclude cases listwise จะได้ดังภาพ




7. Click Continue แล้ว OK จะได้ผลลัพธ์ดังนี้

**ตารางที่ 3.5** ตารางแสดงค่าเฉลี่ยของตัวแปรแต่ละตัวที่ Standardized ใน Cluster ต่าง ๆ

**Initial Cluster Centers**

| | Cluster | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| Zscore: People living in cities (%) | -1.59184 | -1.26130 | 1.63086 | 1.79612 |
| Zscore: Average female life expectancy | -2.47413 | -1.05526 | .74198 | .83657 |
| Zscore: People who read (%) | -2.15601 | -1.15091 | -.23320 | .42230 |
| Zscore: Population increase (% per year)) | .93324 | .18171 | 2.97072 | -.40281 |
| Zscore: Infant mortality (deaths per 1000 live births) | 3.30067 | .96344 | -.78292 | -.96149 |
| Zscore: Birth rate per 1000 people | 2.19054 | .24894 | .16804 | -.80277 |
| Zscore: Death rate per 1000 people | 2.92547 | .10408 | -1.77684 | -.83638 |
| Zscore: Log (base 10) of GDP_CAP | -1.79026 | -1.58451 | .66420 | 1.21600 |
| Zscore: Birth to death ratio | -.37385 | -.14283 | 5.08077 | -.25263 |
| Zscore: Fertility: average number of kids | 1.75399 | .48196 | .22966 | -.88468 |
| Zscore: Log (base 10) of Population | .30228 | 2.82161 | -1.31273 | -.99611 |

จากผลลัพธ์ในตารางเป็นการแสดงค่าเฉลี่ยของตัวแปรแต่ละตัวที่ Standardized ใน Cluster ต่าง ๆ หรือก็คือค่ากลางของ Cluster ในตอนเริ่มต้นนั่นเอง โดยในที่นี้มี 4 กลุ่มหรือ 4 Clusters เนื่องจากได้กำหนดไว้

**ตารางที่ 3.6** แสดงค่าเฉลี่ย หรือค่ากลางของแต่ละ Cluster ที่เปลี่ยนไปในแต่ละรอบของการคำนวณ

**Iteration History**[a]

| Iteration | Change in Cluster Centers | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| 1 | 1.958 | 2.544 | 2.072 | 1.846 |
| 2 | .000 | .442 | .941 | .457 |
| 3 | .193 | .271 | .408 | .187 |
| 4 | .269 | .363 | .357 | .219 |
| 5 | .321 | .336 | .000 | .114 |
| 6 | .000 | .084 | .337 | .000 |
| 7 | .000 | .184 | .501 | .000 |
| 8 | .000 | .182 | .397 | .044 |
| 9 | .000 | .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 9. The minimum distance between initial centers is 5.381.

จากตาราง พบว่า ในตัวอย่างนี้กำหนดให้มีจำนวนรอบสูงสุด = 10 รอบ แต่ในตารางแสดงแค่ 9 รอบ (Iteration) เนื่องจากในรอบที่ 9 ไม่มีการเปลี่ยนแปลงของค่ากลางเมื่อเทียบกับค่ากลางของรอบที่ 8 (ใน Iteration ที่ 9 ค่าที่เปลี่ยนไปของค่ากลางเป็นศูนย์หมด)

**ตารางที่ 3**.7 ตารางแสดงถึง Cluster ที่แต่ละ Case อยู่

**Cluster Membership**

| Case Number | country | Cluster | Distance |
|---|---|---|---|
| 1 | Afghanistan | 1 | 2.468 |
| 2 | Argentina | 4 | 1.573 |
| 3 | Armenia | 4 | 2.051 |
| 4 | Australia | 4 | 1.091 |
| 5 | Austria | 4 | 1.037 |
| 6 | Azerbaijan | 2 | 1.725 |
| 7 | Bahrain | 3 | 2.053 |
| 8 | Bangladesh | 1 | 2.345 |
| 9 | Barbados | 4 | 2.806 |
| 10 | Belarus | 4 | .707 |
| 11 | Belgium | 4 | 1.210 |
| 12 | Bolivia | 2 | 1.507 |
| 13 | Bosnia | . | . |
| 14 | Botswana | 2 | 2.511 |
| 15 | Brazil | 2 | 2.197 |
| 16 | Bulgaria | 4 | 1.184 |
| 17 | Burkina Faso | 1 | 1.295 |
| 18 | Burundi | 1 | 1.734 |
| 19 | Cambodia | 1 | .771 |
| 20 | Cameroon | 1 | 1.862 |
| 21 | Canada | 4 | .975 |
| 22 | Cent. Afri.R | 1 | 2.245 |
| 23 | Chile | 4 | 2.080 |
| 24 | China | 2 | 3.250 |
| 25 | Colombia | 2 | 1.422 |
| 26 | Costa Rica | 3 | 2.436 |
| 27 | Croatia | 4 | 1.375 |
| 28 | Cuba | 4 | 1.549 |
| 29 | Czech Rep. | . | . |
| 30 | Denmark | 4 | 1.178 |
| 31 | Domincan R. | 2 | 1.073 |
| 32 | Ecuador | 2 | .975 |
| 33 | Egypt | 2 | 1.863 |
| 34 | El Salvador | 2 | 1.147 |
| 35 | Estonia | 4 | 1.497 |
| 36 | Ethiopia | 1 | 1.661 |
| 37 | Finland | 4 | .940 |

| Case Number | country | Cluster | Distance |
|---|---|---|---|
| 38 | France | 4 | 1.287 |
| 39 | Gabon | 2 | 3.365 |
| 40 | Gambia | 1 | 1.953 |
| 41 | Georgia | 4 | 1.103 |
| 42 | Germany | 4 | 1.614 |
| 43 | Greece | 4 | .758 |
| 44 | Guatemala | 2 | 1.596 |
| 45 | Haiti | 1 | 1.552 |
| 46 | Honduras | 2 | 1.727 |
| 47 | Hong Kong | 4 | 1.698 |
| 48 | Hungary | 4 | 1.310 |
| 49 | Iceland | 4 | 2.769 |
| 50 | India | 2 | 3.441 |
| 51 | Indonesia | 2 | 2.100 |
| 52 | Iran | 2 | 2.614 |
| 53 | Iraq | 3 | 2.043 |
| 54 | Ireland | 4 | 1.057 |
| 55 | Israel | 4 | 2.046 |
| 56 | Italy | 4 | 1.364 |
| 57 | Japan | 4 | 1.889 |
| 58 | Jordan | 3 | .987 |
| 59 | Kenya | 1 | 2.098 |
| 60 | Kuwait | 3 | 3.776 |
| 61 | Latvia | 4 | 1.223 |
| 62 | Lebanon | 2 | 1.907 |
| 63 | Liberia | 1 | 1.838 |
| 64 | Libya | 3 | 2.026 |
| 65 | Lithuania | 4 | .833 |
| 66 | Malaysia | 2 | 1.366 |
| 67 | Mexico | 2 | 1.961 |
| 68 | Morocco | 2 | 1.315 |
| 69 | N. Korea | 2 | 1.487 |
| 70 | Netherlands | 4 | .905 |
| 71 | New Zealand | 4 | 1.038 |
| 72 | Nicaragua | 2 | 1.970 |
| 73 | Nigeria | 1 | 2.064 |
| 74 | Norway | 4 | .880 |
| 75 | Oman | . | . |
| 76 | Pakistan | 1 | 2.295 |

| Case Number | country | Cluster | Distance |
|---|---|---|---|
| 77 | Panama | 2 | 2.152 |
| 78 | Paraguay | 3 | 1.644 |
| 79 | Peru | 2 | .982 |
| 80 | Philippines | 2 | 1.025 |
| 81 | Poland | 4 | 1.129 |
| 82 | Portugal | 4 | 1.749 |
| 83 | Romania | 4 | 1.402 |
| 84 | Russia | 4 | 1.914 |
| 85 | Rwanda | 1 | 1.898 |
| 86 | S. Korea | 4 | 1.493 |
| 87 | Saudi Arabia | 3 | 1.654 |
| 88 | Senegal | 1 | 1.734 |
| 89 | Singapore | 4 | 1.870 |
| 90 | Somalia | 1 | 1.907 |
| 91 | South Africa | 2 | 1.109 |
| 92 | Spain | 4 | 1.069 |
| 93 | Sweden | 4 | .825 |
| 94 | Switzerland | 4 | .976 |
| 95 | Syria | 3 | 1.718 |
| 96 | Taiwan | . | . |
| 97 | Tanzania | 1 | 1.151 |
| 98 | Thailand | 2 | 2.026 |
| 99 | Turkey | 2 | 1.198 |
| 100 | U.Arab Em. | 3 | 1.949 |
| 101 | UK | 4 | 1.460 |
| 102 | USA | 4 | 2.243 |
| 103 | Uganda | 1 | 2.228 |
| 104 | Ukraine | 4 | 1.794 |
| 105 | Uruguay | 4 | 1.375 |
| 106 | Uzbekistan | 2 | 1.011 |
| 107 | Venezuela | 2 | 2.167 |
| 108 | Vietnam | 2 | 2.015 |
| 109 | Zambia | 1 | 1.886 |

จากตารางแสดงถึง Cluster ที่แต่ละ case อยู่เช่น case ที่ 1 คือ ประเทศ Afghanistan อยู่ใน Cluster ที่ 1 และมีระยะห่างจากค่ากลางของ Cluster ที่ 1 มากที่สุดคือ 2.468 ,....., case ที่ 109 คือ ประเทศ Zambia อยู่ใน Cluster ที่ 1 และมีระยะห่างจากค่ากลางของ Cluster ที่ 1 คือ 1.886

**ตารางที่ 3.8** ตารางแสดงค่าเฉลี่ยตัวแปรที่ Standardized แล้ว

**Final Cluster Centers**

| | Cluster | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| Zscore: People living in cities (%) | -1.30882 | -.25371 | .57315 | .69184 |
| Zscore: Average female life expectancy | -1.79780 | -.07883 | .22173 | .77423 |
| Zscore: People who read (%) | -1.61850 | -.04713 | -.22883 | .80070 |
| Zscore: Population increase (% per year)) | .90777 | .30535 | 1.52110 | -.94615 |
| Zscore: Infant mortality (deaths per 1000 live births) | 1.71845 | .17485 | -.17970 | -.80052 |
| Zscore: Birth rate per 1000 people | 1.50289 | .18108 | .76670 | -.94618 |
| Zscore: Death rate per 1000 people | 1.53829 | -.55197 | -1.08325 | -.03111 |
| Zscore: Log (base 10) of GDP_CAP | -1.38113 | -.49446 | .23922 | .87086 |
| Zscore: Birth to death ratio | -.12959 | .41270 | 2.15339 | -.74000 |
| Zscore: Fertility: average number of kids | 1.49223 | -.01451 | .85674 | -.87763 |
| Zscore: Log (base 10) of Population | .03887 | .40569 | -.66643 | -.11116 |

จากผลลัพธ์ในตาราง เป็นค่าเฉลี่ยตัวแปรที่ Standardized แล้ว ค่าเฉลี่ยเหล่านี้คือ ค่ากลางของแต่ละ Cluster จะพบว่าค่าเฉลี่ยของตัวแปร babymort จะแตกต่างกันเมื่ออยู่ Cluster ที่ต่างกัน และแตกต่างกันมากเมื่อเทียบกับตัวแปรอื่นๆ นั่นคือค่าเฉลี่ยของ babymort ใน Cluster ที่ 1=1.71845 หรือ มากกว่าค่าเฉลี่ยรวม 1.71845 เท่าของค่าเบี่ยงเบนมาตรฐาน ขณะที่ของ Cluster ที่ 4 เป็น -.80052 หรือน้อยกว่าค่าเฉลี่ยรวมถึง .8 เท่าของค่าเบี่ยงเบนมาตรฐาน ใน ทำนองเดียวกับตัวแปร lifeexp, birth_literac ก็มีค่าเฉลี่ยแตกต่างกันมากเมื่ออยู่ต่าง Cluster กัน

**ตารางที่ 3.9** ตารางแสดงระยะห่างของค่ากลางของทั้ง 4 Cluster

**Distances between Final Cluster Centers**

| Cluster | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | | 4.346 | 5.448 | 6.767 |
| 2 | 4.346 | | 2.918 | 3.257 |
| 3 | 5.448 | 2.918 | | 4.897 |
| 4 | 6.767 | 3.257 | 4.897 | |

จากตาราง พบว่า Cluster ที่ 1 มีระยะห่างจาก Cluster ที่ 4 มากที่สุด คือ 6.767 และอยู่ใกล้ Cluster ที่ 2 มากที่สุด คือ 4.346 ส่วน Cluster ที่ 3 อยู่ใกล้ Cluster ที่ 2 และ Cluster ที่ 4 อยู่ใกล้ Cluster ที่ 2

**ตารางที่ 3.10** ตารางแสดงค่า Mean Square ระหว่าง Cluster และ Mean Square Error

ANOVA

| | Cluster | | Error | | F | Sig. |
|---|---|---|---|---|---|---|
| | Mean Square | df | Mean Square | df | | |
| Zscore: People living in cities (%) | 20.186 | 3 | .414 | 101 | 48.794 | .000 |
| Zscore: Average female life expectancy | 30.553 | 3 | .146 | 101 | 208.801 | .000 |
| Zscore: People who read (%) | 27.062 | 3 | .241 | 101 | 112.070 | .000 |
| Zscore: Population increase (% per year)) | 27.295 | 3 | .211 | 101 | 129.369 | .000 |
| Zscore: Infant mortality (deaths per 1000 live births) | 29.485 | 3 | .170 | 101 | 173.760 | .000 |
| Zscore: Birth rate per 1000 people | 30.477 | 3 | .124 | 101 | 245.816 | .000 |
| Zscore: Death rate per 1000 people | 22.843 | 3 | .363 | 101 | 63.005 | .000 |
| Zscore: Log (base 10) of GDP_CAP | 26.541 | 3 | .265 | 101 | 100.007 | .000 |
| Zscore: Birth to death ratio | 25.358 | 3 | .244 | 101 | 103.766 | .000 |
| Zscore: Fertility: average number of kids | 28.589 | 3 | .168 | 101 | 170.097 | .000 |
| Zscore: Log (base 10) of Population | 3.362 | 3 | .947 | 101 | 3.550 | .017 |

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.

จากตาราง พบว่า ค่าเฉลี่ยของตัวแปร Birth rate per 1000 people เมื่ออยู่ต่างกลุ่มกันจะมีความแตกต่างกันมากที่สุด เนื่องจากมีค่าสถิติทดสอบ F สูงที่สุด คือ 245.816 และรองลงมา คือ ตัวแปร Average female life expectancy ซึ่งมีค่าสถิติทดสอบ F คือ 208.801

**ตารางที่ 3.11** ตารางแสดงจำนวน Case หรือประเทศที่อยู่ในแต่ละ Cluster

**Number of Cases in each Cluster**

| | | |
|---|---|---|
| Cluster | 1 | 20.000 |
| | 2 | 31.000 |
| | 3 | 10.000 |
| | 4 | 44.000 |
| Valid | | 105.000 |
| Missing | | 4.000 |

จากตารางแสดงจำนวน Case หรือประเทศที่อยู่ในแต่ละ Cluster พบว่า ประเทศส่วนใหญ่จะอยู่ใน Cluster ที่ 4 คือมีจำนวน 44 ประเทศ และ Cluster ที่ 3 จะมีจำนวนประเทศน้อยที่สุด คือมีจำนวน 10 ประเทศ

**4. บทสรุป**

Cluster Analysis (CA) เป็นเทคนิคที่ใช้จำแนกหรือแบ่งหน่วยสังเกตหรือหน่วยที่ให้ข้อมูล (Case) ออกเป็นกลุ่มย่อย ๆ ตั้งแต่ 2 กลุ่มขึ้นไป โดยหลักการของ CA ก็คือ การจัดกลุ่มให้กับ Case ที่อยู่ในกลุ่มเดียวกันมีลักษณะที่เหมือนกันหรือคล้ายคลึงกันให้ได้มากที่สุด ส่วน Case ที่อยู่ต่างกลุ่มกันมีลักษณะที่แตกต่างกันให้ได้มากที่สุด โดยเทคนิคที่ใช้กันมากมี 2 เทคนิค คือ 1.เทคนิค Hierarchical Cluster Analysis ซึ่งไม่จำเป็นต้องระบุจำนวนกลุ่ม และเหมาะกับกลุ่มตัวอย่างขนาดเล็ก และ 2. เทคนิค K-Means Cluster Analysis ซึ่งผู้วิเคราะห์จะต้องกำหนดจำนวนกลุ่มที่แน่นอนไว้ล่วงหน้า และต้องมีการ Standardized ข้อมูลก่อนการวิเคราะห์ โดยเทคนิคนี้เหมาะกับกลุ่มตัวอย่างที่มีมากกว่าหรือเท่ากับ 200 ตัวอย่าง

# แบบฝึกหัดบทที่ 3
# เรื่อง การจำแนกกลุ่มตัวแปรด้วยเทคนิค Cluster Analysis

---

1. นักวิจัยผู้หนึ่งได้สำรวจบริษัทขนาดเล็ก เกี่ยวกับความสนใจที่จะส่งสินค้าไปจำหน่ายต่างประเทศ โดยสุ่มตัวอย่างบริษัทขนาดเล็กมา 30 บริษัท และส่งแบบสอบถามเกี่ยวกับระดับความสนใจที่จะส่งสินค้าไปขายต่างประเทศ นำคำตอบที่ได้มาสร้างแฟ้มข้อมูลชื่อ export.sav ซึ่งประกอบด้วยตัวแปรดังนี้

| ชื่อตัวแปร | ความหมาย | Measure | ชนิด (Type) | ค่าของตัวแปร |
|---|---|---|---|---|
| Will | ระดับความสนใจที่จะส่ง | Ordinal | Numeric | 1 = ไม่สนใจส่งออกเลย<br>2 = ไม่สนใจส่งออก<br>3 = เฉยๆ<br>4 = สนใจส่งออก<br>5 = สนใจส่งออกอย่างมาก |
| Size | จำนวนพนักงาน | Scale | Numeric | ระบุจำนวนพนักงานตามจริง |
| Revenue | รายได้ต่อปีของบริษัท | Scale | Numeric | ระบุรายได้ตามจริง : ล้านบาท |
| Product | จำนวนผลิตภัณฑ์ที่ผลิต | Scale | Numeric | ระบุจำนวนผลิตภัณฑ์ตามจริง : ล้านชิ้น |
| Years | ระยะเวลาที่ขายในประเทศ | Scale | Numeric | ระบุระยะเวลาตามจริง |

ได้ข้อมูลดังตาราง

| ลำดับที่ | Will | Size | Revenue | Product | Years |
|---|---|---|---|---|---|
| 1 | 5 | 54 | 3500 | 5 | 1 |
| 2 | 5 | 40 | 2100 | 10 | 2 |
| 3 | 3 | 60 | 3700 | 12 | 4 |
| 4 | 4 | 56 | 4800 | 4 | 3 |
| 5 | 1 | 46 | 5600 | 5 | 2 |
| 6 | 2 | 67 | 3600 | 15 | 1 |
| 7 | 3 | 70 | 2200 | 5 | 4 |
| 8 | 4 | 35 | 3500 | 6 | 5 |
| 9 | 5 | 69 | 3800 | 5 | 5 |
| 10 | 3 | 75 | 2570 | 7 | 6 |
| 11 | 3 | 80 | 4500 | 14 | 3 |
| 12 | 4 | 55 | 4800 | 5 | 4 |
| 13 | 5 | 46 | 2200 | 3 | 2 |
| 14 | 1 | 50 | 4500 | 2 | 5 |

| ลำดับที่ | Will | Size | Revenue | Product | Years |
|---|---|---|---|---|---|
| 15 | 2 | 73 | 5500 | 4 | 3 |
| 16 | 3 | 47 | 4900 | 1 | 2 |
| 17 | 4 | 58 | 3200 | 4 | 1 |
| 18 | 4 | 62 | 2300 | 5 | 4 |
| 19 | 5 | 84 | 5400 | 20 | 3 |
| 20 | 5 | 47 | 5700 | 6 | 2 |
| 21 | 2 | 55 | 3800 | 2 | 1 |
| 22 | 3 | 80 | 4500 | 1 | 4 |
| 23 | 1 | 72 | 3200 | 3 | 5 |
| 24 | 1 | 89 | 5800 | 7 | 3 |
| 25 | 2 | 52 | 2200 | 5 | 2 |
| 26 | 3 | 90 | 3400 | 24 | 1 |
| 27 | 4 | 48 | 5600 | 5 | 2 |
| 28 | 5 | 65 | 6700 | 7 | 3 |
| 29 | 3 | 55 | 5500 | 3 | 4 |
| 30 | 5 | 86 | 5600 | 22 | 1 |

จงจัดกลุ่มตัวแปรโดยใช้เทคนิค Hierarchical Cluster Analysis พร้อมทั้งอธิบายผลลัพธ์ที่ได้จากการวิเคราะห์มาโดยละเอียด

2. จากไฟล์แบบฝึกหัดบทที่ 3 ข้อ 2 ซึ่งเป็นข้อมูลของบริษัทสื่อสารแห่งหนึ่งผู้วิเคราะห์ต้องการแบ่งกลุ่มลูกค้าตามรูปแบบการใช้บริการ เพื่อนำไปใช้เป็นข้อมูลในการออกแบบผลิตภัณฑ์ที่ตรงตามความต้องการของลูกค้าแต่ละกลุ่ม จงจัดกลุ่มตัวแปรโดยใช้เทคนิค K - Mean Cluster Analysis โดยเลือกตัวแปร tenure ,age ,income ,reside ,longmon ,tollmon ,eguipmon ,cardmon ,wiremon มาใช้ในการวิเคราะห์โดยทำการแบ่งกลุ่ม Cluster เป็น 3 Cluster และไม่ต้องกำหนด Label case by พร้อมทั้งอธิบายผลลัพธ์ที่ได้จากการวิเคราะห์มาโดยละเอียด

# เอกสารอ้างอิงบทที่ 3

กัลยา วานิชย์บัญชา. (2551). **การวิเคราะห์สถิติขั้นสูงด้วย SPSS for Windows**. พิมพ์ครั้งที่ 6. กรุงเทพฯ: โรงพิมพ์จุฬาลงกรณ์มหาวิทยาลัย, จุฬาลงกรณ์มหาวิทยาลัย.

บุญชม ศรีสะอาด. (2543). **การวิจัยทางการวัดผลและประเมินผล.** กรุงเทพฯ : สุวีริยาสาส์น.

ศิริชัย พงษ์วิชัย. (2544). **การวิเคราะห์ข้อมูลทางสถิติด้วยคอมพิวเตอร์.** พิมพ์ครั้งที่ 11. กรุงเทพฯ : จุฬาลงกรณ์มหาวิทยาลัย.

สุชาติ ประสิทธิ์รัฐสินธุ์. 2540. **เทคนิคการวิเคราะห์ตัวแปรหลายตัวสำหรับการวิจัยทางสังคมศาสตร์และพฤติกรรมศาสตร์.** พิมพ์ครั้งที่ 4. กรุงเทพฯ : เลี่ยงเชียง.

**เทคนิคการจัดกลุ่ม (Clustering).** (2557). ค้นเมื่อ 21 เมษายน 2557, จาก http://datamining-techniques.blogspot.com/2012/09/clustering.html.

**การจำแนกกลุ่มตัวแปรด้วยเทคนิค Cluster Analysis.** (2557). ค้นเมื่อ 22 เมษายน 2557, จาก http://www.saruthipong.com/port/document/299-705/299-705-10.pdf.