การเปรียบเทียบวิธีการจัดกลุ่มสำหรับข้อมูลที่มีการแจกแจงปกติแบบผสม

นางสาวจิรวรรณ ไพบูลย์วรชาติ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาสถิติ ภาควิชาสถิติ
คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2556
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

บทคัดย่อและแฟ้มข้อมูลฉบับเต็มของวิทยานิพนธ์ตั้งแต่ปีการศึกษา 2554 ที่ให้บริการในคลังปัญญาจุฬาฯ (CUIR)
เป็นแฟ้มข้อมูลของนิสิตเจ้าของวิทยานิพนธ์ ที่ส่งผ่านทางบัณฑิตวิทยาลัย
The abstract and full text of theses from the academic year 2011 in Chulalongkorn University Intellectual Repository (CUIR)
are the thesis authors' files submitted through the University Graduate School.

COMPARISON OF CLUSTERING ALGORITHMS FOR MIXTURES OF GAUSSIAN DISTRIBUTION

Miss Jirawun Paiboolworachat

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science Program in Statistics
Department of Statistics
Faculty of Commerce and Accountancy
Chulalongkorn University
Academic Year 2013

Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์	การเปรียบเทียบวิธีการจัดกลุ่มสำหรับข้อมูลที่มีการแจกแจงปกติแบบผสม
โดย	นางสาวจิรวรรณ ไพบูลย์วรชาติ
สาขาวิชา	สถิติ
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	อาจารย์ ดร. นัท กุลวานิช

คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

.....คณบดีคณะพาณิชยศาสตร์และการบัญชี

(รองศาสตราจารย์ ดร. พสุ เดชะรินทร์)

คณะกรรมการสอบวิทยานิพนธ์

.....ประธานกรรมการ

(รองศาสตราจารย์ ดร. เสกสรร เกียรติสุไพบูลย์)

.....อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

(อาจารย์ ดร. นัท กุลวานิช)

.....กรรมการ

(อาจารย์ ดร. อัครินทร์ ไพบูลย์พานิช)

.....กรรมการภายนอกมหาวิทยาลัย

(อาจารย์ ดร. อรุณี กำลัง)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จิรวรรณ ไพบูลย์วรชาติ : การเปรียบเทียบวิธีการจัดกลุ่มสำหรับข้อมูลที่มีการแจกแจงปกติแบบผสม. (COMPARISON OF CLUSTERING ALGORITHMS FOR MIXTURES OF GAUSSIAN DISTRIBUTION) อ.ที่ปรึกษาวิทยานิพนธ์หลัก: อ. ดร. นัท กุลวานิช, 126 หน้า.

งานวิจัยนี้จึงมีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพวิธีการจัดกลุ่มข้อมูล 4 วิธี คือ วิธีการจัดกลุ่มแบบลำดับชั้น, วิธีการจัดกลุ่มแบบเค-มีน, วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน และ วิธีการจัดกลุ่มแบบอัลกอริทึม EM โดยทำการจำลองข้อมูลที่มีการแจกแจงปกติแบบผสม ซึ่งแบ่งออกเป็น 2 กรณี ดังนี้ 1.กรณีที่ฐานข้อมูลอยู่ในรูปแบบวงรี (Non Spherical) 2.กรณีที่ฐานข้อมูลอยู่ในรูปแบบวงกลม (Spherical or Isotopic) ทำการจำลองข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 2, 3 และ 4 กลุ่ม ตัวแปรที่ใช้ในการศึกษามี 2 และ 3 ตัวแปร ขนาดข้อมูลแต่ละกลุ่มเท่ากับ 50, 100 และ 300 ตัวอย่าง กำหนดจำนวนกลุ่มของการจัดกลุ่มข้อมูลเท่ากับ 2, 3 และ 4 กลุ่ม ทำการเปรียบเทียบประสิทธิภาพของวิธีการจัดกลุ่มข้อมูลจาก 2 วิธี คือ วิธี Calinski and Harabasz index (Pseudo F) และวิธี Silhouette width พบว่า เมื่อทำการจำลองข้อมูลกรณีที่ฐานข้อมูลอยู่ในรูปแบบวงรี (Non Spherical) วิธีการจัดกลุ่มข้อมูลทั้ง 4 วิธี เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพที่ดี ตามแต่ละสถานการณ์ เมื่อทำการจำลองข้อมูลกรณีที่ฐานข้อมูลอยู่ในรูปแบบวงกลม (Spherical or Isotopic) วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพที่ดี เมื่อจำนวนกลุ่มของการซ้อนทับ และอัตราการซ้อนทับเฉลี่ยมีค่าเพิ่มขึ้น

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ภาควิชา สถิติ ลายมือชื่อนิสิต .....

สาขาวิชา สถิติ ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์หลัก .....

ปีการศึกษา 2556

5581518026 : MAJOR STATISTICS
KEYWORDS: CLUSTERING / GAUSSIAN DISTRIBUTION

JIRAWUN PAIBOOLWORACHAT: COMPARISON OF CLUSTERING ALGORITHMS FOR MIXTURES OF GAUSSIAN DISTRIBUTION. ADVISOR: NAT KULVANICH, Ph.D., 126 pp.

The purpose of this research is to compare the efficiency of 4 clustering. 4 clustering are Hierarchical Clustering, K-Means Clustering, Fuzzy C-Means Clustering and Expectation-Maximization Algorithm (EM Clustering). The simulated data with mixture of Gaussian distribution can be considered into 2 cases which are Non-Spherical and Spherical. The simulations of the data with overlap 2, 3 and 4 clusters have 2 and 3 variables and the sample size of each data is 50, 100, and 300. There are 2 clustering which are Calinski and Harabasz index (Pseudo F) and Silhouette width. When the database is simulated in Non spherical, the 4 methods clustering are the effective methods base on situation. However, when the database is simulated in spherical, EM is the most effective method which depends on the increasing of overlapping number and the average of overlap.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

Department: Statistics
Field of Study: Statistics
Academic Year: 2013

Student's Signature .....

Advisor's Signature .....

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สามารถสำเร็จลุล่วงไปได้ด้วยดี เนื่องจากผู้วิจัยได้รับความช่วยเหลือ และเอาใจใส่อย่างดียิ่งของอาจารย์ ดร. นัท กุลวานิช อาจารย์ที่ปรึกษาวิทยานิพนธ์ ได้กรุณาสละเวลาให้ความช่วยเหลือ คำปรึกษา ข้อชี้แนะ ข้อคิดเห็น ผู้วิจัยขอขอบพระคุณเป็นอย่างยิ่ง

ผู้วิจัยขอขอบพระคุณ รองศาสตราจารย์ ดร.เสกสรร เกียรติสุไพบูลย์ ประธานกรรมการสอบวิทยานิพนธ์ อาจารย์ ดร.อัครินทร์ ไพบูลย์พานิช และ อาจารย์ ดร.อรุณี กำลัง กรรมการสอบวิทยานิพนธ์ ที่กรุณาให้คำแนะนำ แนวคิดที่เป็นประโยชน์ รวมถึง การตรวจสอบ และแก้ไขวิทยานิพนธ์ฉบับนี้ให้สมบูรณ์มากยิ่งขึ้น

ผู้วิจัยขอขอบพระคุณ คณาจารย์ทุกท่าน ของ ภาคสถิติ คณะพาณิชยศาสตร์ และ การบัญชี จุฬาลงกรณ์มหาวิทยาลัย ที่ถ่ายทอดความรู้ความสามารถทางสถิติให้กับผู้วิจัย

ผู้วิจัยขอขอบคุณ เจ้าหน้าที่ทุกท่าน ของ ภาคสถิติ คณะพาณิชยศาสตร์ และ การบัญชี จุฬาลงกรณ์มหาวิทยาลัย ที่อำนวยความสะดวกต่างๆในเรื่องการทำวิจัย และขอบคุณเพื่อนๆนิสิตภาควิชาสถิติ คณะพาณิชยศาสตร์ และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย ที่ให้กำลังใจ และความช่วยเหลือโดยตลอด

สุดท้ายนี้ ผู้วิจัยขอกราบขอบพระคุณครอบครัว ที่ช่วยส่งเสริม ให้กำลังใจ และ สนับสนุนความสำเร็จของผู้วิจัย จนทำให้ผู้วิจัยสามารถประสบความสำเร็จไปด้วยดีอยู่เสมอ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ง
บทคัดย่อภาษาอังกฤษ ..... จ
กิตติกรรมประกาศ ..... ฉ
สารบัญ ..... ช
สารบัญตาราง ..... ฌ
สารบัญภาพ ..... ด
บทที่ 1 บทนำ 1
1.1 ความเป็นมาและความสำคัญของปัญหา 1
1.2 วัตถุประสงค์ 2
1.3 ข้อตกลงเบื้องต้น 2
1.4 ขอบเขตของการวิจัย 2
1.5 วิธีดำเนินการวิจัย 3
1.6 ประโยชน์ที่คาดว่าจะได้รับ 4
บทที่ 2 ทฤษฎีและตัวสถิติที่เกี่ยวข้อง 5
2.1 การแจกแจงปกติหลายตัวแปร (Multivariate Normal distribution) 5
2.2 เมทริกซ์ความแปรปรวนร่วม (Covariance Matrix) 5
2.3 แพคเกจ MixSim 5
2.4 การจัดกลุ่มแบบลำดับชั้น (Hierarchical Clustering) 6
2.5 วิธีการจัดกลุ่มแบบเคมีน (K-Means Clustering) 13
2.6 วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน (Fuzzy C-Means Clustering) 17
2.7 เทคนิคการจัดกลุ่มข้อมูลแบบอัลกอริทึม EM (Expectation-Maximization Algorithm) 23
2.8 การวัดประสิทธิภาพการจัดกลุ่มข้อมูล 27
2.8.1 วิธี Calinski and Harabasz index (Pseudo F) 27
2.8.2 วิธี Silhouette width 28
บทที่ 3 วิธีการดำเนินงานวิจัย 29
3.1 ขอบเขตของการดำเนินการวิจัย 29
3.2 ขั้นตอนในการดำเนินการวิจัย 29

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

หน้า

3.3 ขั้นตอนการทำงานของโปรแกรม 31
บทที่ 4 ผลการวิเคราะห์ข้อมูล 32
4.1 กรณีที่ฐานข้อมูลอยู่ในรูปแบบวงรี (Non Spherical) 32
4.1.1 วิธี Calinski and Harabasz index (Pseudo F) 32
4.1.2 วิธี Silhouette width 52
4.2 กรณีที่ฐานข้อมูลอยู่ในรูปแบบวงกลม (Spherical or Isotopic) 70
4.2.1 วิธี Calinski and Harabasz index (Pseudo F) 70
4.2.2 วิธี Silhouette width 89
4.3 ข้อมูลจริง 107
บทที่ 5 สรุปผลงานวิจัย อภิปรายผล และข้อเสนอแนะ 111
5.1 สรุปผลงานวิจัย อภิปรายผล 112
5.2 ข้อเสนอแนะ 122
รายการอ้างอิง 123
ภาคผนวก 124
ประวัติผู้เขียนวิทยานิพนธ์ 126

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

สารบัญตาราง

ตารางที่ หน้า

กรณีที่ฐานข้อมูลอยู่ในรูปแบบวงรี (Non Spherical)

4.1.1 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม ขนาดตัวอย่างกลุ่มละ 50 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร 33

4.1.2 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม ขนาดตัวอย่างกลุ่มละ 100 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร 34

4.1.3 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม ขนาดตัวอย่างกลุ่มละ 300 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร 35

4.1.4 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม ขนาดตัวอย่างกลุ่มละ 50 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร 36

4.1.5 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม ขนาดตัวอย่างกลุ่มละ 100 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร 38

4.1.6 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม ขนาดตัวอย่างกลุ่มละ 300 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร 39

4.1.7 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม ขนาดตัวอย่างกลุ่มละ 50 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร 40

4.1.8 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม ขนาดตัวอย่างกลุ่มละ 100 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร 41

4.1.9 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม ขนาดตัวอย่างกลุ่มละ 300 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร 42

4.1.10 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม ขนาดตัวอย่างกลุ่มละ 50 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร 43

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ หน้า

4.1.11 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม ขนาดตัวอย่างกลุ่มละ 100 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร 44

4.1.12 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม ขนาดตัวอย่างกลุ่มละ 300 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร 45

4.1.13 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม ขนาดตัวอย่างกลุ่มละ 50 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร 46

4.1.14 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม ขนาดตัวอย่างกลุ่มละ 100 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร 47

4.1.15 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม ขนาดตัวอย่างกลุ่มละ 300 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร 48

4.1.16 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม ขนาดตัวอย่างกลุ่มละ 50 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร 49

4.1.17 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม ขนาดตัวอย่างกลุ่มละ 100 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร 50

4.1.18 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม ขนาดตัวอย่างกลุ่มละ 300 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร 51

4.1.19 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม ขนาดตัวอย่างกลุ่มละ 50 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร 52

4.1.20 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม ขนาดตัวอย่างกลุ่มละ 100 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษามีความสัมพันธ์กัน 2 ตัวแปร 53

4.1.21 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม ขนาดตัวอย่างกลุ่มละ 300 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร 54

ตารางที่ หน้า

4.1.22 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม ขนาดตัวอย่างกลุ่มละ 50 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร 55
4.1.23 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม ขนาดตัวอย่างกลุ่มละ 100 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร 56
4.1.24 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม ขนาดตัวอย่างกลุ่มละ 300 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร 57
4.1.25 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม ขนาดตัวอย่างกลุ่มละ 50 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร 58
4.1.26 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม ขนาดตัวอย่างกลุ่มละ 100 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร 59
4.1.27 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม ขนาดตัวอย่างกลุ่มละ 300 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร 60
4.1.28 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม ขนาดตัวอย่างกลุ่มละ 50 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร 61
4.1.29 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม ขนาดตัวอย่างกลุ่มละ 100 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร 62
4.1.30 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม ขนาดตัวอย่างกลุ่มละ 300 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร 63
4.1.31 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม ขนาดตัวอย่างกลุ่มละ 50 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร 64
4.1.32 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม ขนาดตัวอย่างกลุ่มละ 100 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร 65

ตารางที่ หน้า

4.1.33 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม ขนาดตัวอย่างกลุ่มละ 300 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร 66

4.1.34 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม ขนาดตัวอย่างกลุ่มละ 50 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร 67

4.1.35 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม ขนาดตัวอย่างกลุ่มละ 100 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร 68

4.1.36 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม ขนาดตัวอย่างกลุ่มละ 300 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร 69

กรณีที่ฐานข้อมูลอยู่ในรูปแบบวงกลม (Spherical or Isotopic)

4.2.1 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม ขนาดตัวอย่างกลุ่มละ 50 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร 70

4.2.2 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม ขนาดตัวอย่างกลุ่มละ 100 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร 72

4.2.3 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม ขนาดตัวอย่างกลุ่มละ 300 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร 73

4.2.4 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม ขนาดตัวอย่างกลุ่มละ 50 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร 74

4.2.5 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม ขนาดตัวอย่างกลุ่มละ 100 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร 75

4.2.6 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม ขนาดตัวอย่างกลุ่มละ 300 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร 76

ตารางที่ หน้า

4.2.7 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม ขนาดตัวอย่างกลุ่มละ 50 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร 77

4.2.8 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม ขนาดตัวอย่างกลุ่มละ 100 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร 78

4.2.9 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม ขนาดตัวอย่างกลุ่มละ 300 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร 79

4.2.10 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม ขนาดตัวอย่างกลุ่มละ 50 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร 80

4.2.11 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม ขนาดตัวอย่างกลุ่มละ 100 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร 81

4.2.12 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม ขนาดตัวอย่างกลุ่มละ 300 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร 82

4.2.13 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม ขนาดตัวอย่างกลุ่มละ 50 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร 83

4.2.14 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม ขนาดตัวอย่างกลุ่มละ 100 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร 84

4.2.15 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม ขนาดตัวอย่างกลุ่มละ 300 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร 85

4.2.16 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม ขนาดตัวอย่างกลุ่มละ 50 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร 86

4.2.17 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม ขนาดตัวอย่างกลุ่มละ 100 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร 87

ตารางที่ หน้า

4.2.18 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม ขนาดตัวอย่างกลุ่มละ 300 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร 88
4.2.19 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม ขนาดตัวอย่างกลุ่มละ 50 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร 89
4.2.20 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม ขนาดตัวอย่างกลุ่มละ 100 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร 90
4.2.21 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม ขนาดตัวอย่างกลุ่มละ 300 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร 91
4.2.22 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม ขนาดตัวอย่างกลุ่มละ 50 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร 92
4.2.23 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette widthข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม ขนาดตัวอย่างกลุ่มละ 100 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร 93
4.2.24 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม ขนาดตัวอย่างกลุ่มละ 300 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร 94
4.2.25 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม ขนาดตัวอย่างกลุ่มละ 50 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร 95
4.2.26 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม ขนาดตัวอย่างกลุ่มละ 100 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร 96
4.2.27 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม ขนาดตัวอย่างกลุ่มละ 300 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร 97
4.2.28 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม ขนาดตัวอย่างกลุ่มละ 50 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร 98

ตารางที่ หน้า

4.2.29 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม ขนาดตัวอย่างกลุ่มละ 100 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร 99
4.2.30 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม ขนาดตัวอย่างกลุ่มละ 300 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร 100
4.2.31 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม ขนาดตัวอย่างกลุ่มละ 50 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร 101
4.2.32 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม ขนาดตัวอย่างกลุ่มละ 100 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร 102
4.2.33 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม ขนาดตัวอย่างกลุ่มละ 300 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร 103
4.2.34 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม ขนาดตัวอย่างกลุ่มละ 50 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร 104
4.2.35 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม ขนาดตัวอย่างกลุ่มละ 100 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร 105
4.2.36 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม ขนาดตัวอย่างกลุ่มละ 300 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร 106

ข้อมูลจริง

4.3.1 ค่า Pseudo F โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม ขนาดตัวอย่างกลุ่มละ 100 ตัวอย่าง 109
4.3.2 ค่า Silhouette width โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม ขนาดตัวอย่างกลุ่มละ 100 ตัวอย่าง 110
5.1.1 สรุปผลวิธีการจัดกลุ่มข้อมูลด้วยค่า Pseudo F และค่า Silhouette width ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม จำนวนตัวแปรที่ใช้ศึกษา 2 ตัวแปร 112

Chulalongkorn University

ตารางที่ หน้า

5.1.2 สรุปผลวิธีการจัดกลุ่มข้อมูลด้วยค่า Pseudo F และค่า Silhouette width ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม จำนวนตัวแปรที่ใช้ศึกษา 3 ตัวแปร 114

5.1.3 สรุปผลวิธีการจัดกลุ่มข้อมูลด้วยค่า Pseudo F และค่า Silhouette width ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม จำนวนตัวแปรที่ใช้ศึกษา 2 ตัวแปร 116

5.1.4 สรุปผลวิธีการจัดกลุ่มข้อมูลด้วยค่า Pseudo F และค่า Silhouette width ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม จำนวนตัวแปรที่ใช้ศึกษา 3 ตัวแปร 118

5.1.5 สรุปผลวิธีการจัดกลุ่มข้อมูลด้วยค่า Pseudo F และค่า Silhouette width ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม จำนวนตัวแปรที่ใช้ศึกษา 2 ตัวแปร 119

5.1.6 สรุปผลวิธีการจัดกลุ่มข้อมูลด้วยค่า Pseudo F และค่า Silhouette width ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม จำนวนตัวแปรที่ใช้ศึกษา 3 ตัวแปร 121

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

สารบัญภาพ

ภาพที่		หน้า
2.3.1	ขั้นตอนในการจำลองข้อมูล	6
2.4.1	ขั้นตอนการจัดกลุ่มแบบลำดับขั้นประเภทการเกาะกลุ่ม	8
2.4.2	กราฟเดนโดรแกรม (Dendrogram)	13
2.5.1	ขั้นตอนการจัดกลุ่มแบบเคมีน	14
2.6.1	ขั้นตอนการจัดกลุ่มแบบฟัซซี่ซีมีน	19
2.7.1	ขั้นตอนการจัดกลุ่มข้อมูลแบบอัลกอริทึม EM	26
3.3	ผังการทำงานของโปรแกรม	31
4.3.1	กราฟ Chi-square ของข้อมูลตามเพศ	108

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

การแบ่งกลุ่มข้อมูลเป็นเทคนิคที่ใช้แบ่งกลุ่ม คน สัตว์ สิ่งของ ฯลฯ ที่มีลักษณะเหมือนหรือคล้ายกันไว้ในกลุ่มเดียวกัน ซึ่งมีหลายเทคนิคที่ใช้ในการพิจารณาความคล้ายกัน เช่น เทคนิคการวิเคราะห์กลุ่ม (Cluster Analysis), เทคนิคการวิเคราะห์จำแนกกลุ่ม (Discriminant Analysis), เทคนิคการวิเคราะห์ความถดถอยโลจิสติก (Logistic Regression Analysis) เป็นต้น เทคนิคการวิเคราะห์หรือการจัดกลุ่มข้อมูลเหล่านี้ เป็นเทคนิคที่นิยมใช้กันอย่างแพร่หลาย เนื่องจากมีความแตกต่างจากเทคนิคอื่น ๆ ตรงที่ไม่จำเป็นต้องทราบจำนวนกลุ่มของข้อมูลมาก่อน หรือระบุจำนวนกลุ่มของข้อมูลไม่ได้ (กัลยา วานิชย์บัญชา, 2552) สำหรับวิธีการจัดกลุ่มข้อมูลที่จะกล่าวถึงมีดังนี้

วิธีการจัดกลุ่มแบบลำดับชั้น (Hierarchical Clustering) นำเสนอโดย Johnson (1967) ป็นวิธีการจัดกลุ่มที่นิยมนำไปใช้ โดยเมื่อทำการจัดกลุ่มข้อมูลแล้ว ข้อมูลตัวอย่างเหล่านั้นจะไม่ถูกย้ายหรือเปลี่ยนไปยังกลุ่มอื่น ๆ อีก หลังจากนั้น MacQueen (1967) ได้นำเสนอวิธีการจัดกลุ่มแบบเคมีน (K-Means Clustering) เป็นวิธีการจัดกลุ่มแบบแบ่งส่วน (Partitioning) ซึ่งสามารถใช้งานได้ดีกว่าการจัดกลุ่มแบบลำดับชั้น เมื่อข้อมูลมีขนาดใหญ่ ส่วน Dunn (1973) ได้คิดค้นวิธีการจัดกลุ่มแบบฟัซซี่ซีมีน (Fuzzy C Means Clustering) และ Bezdek (1981) ได้นำไปพัฒนาต่อ โดยเมื่อทำการจัดกลุ่มข้อมูลแบบฟัซซี่ซีมีนแล้ว หากข้อมูลตัวอย่างเหล่านั้นมีความเกี่ยวข้องกับกลุ่มอื่นด้วย วิธีนี้จะแสดงความน่าจะเป็นที่ข้อมูลจะถูกจัดเข้าในแต่ละกลุ่ม ซึ่งมักถูกใช้ในการจดจำรูปแบบ หรือการจดจำรูปภาพ หลังจากนั้น Dempster, Laird et al. (1977) ได้นำเสนอวิธีการจัดกลุ่มแบบอัลกอริทึม EM (Expectation-Maximization Algorithm; EM Clustering) ซึ่งเป็นวิธีการจัดกลุ่มข้อมูลที่แต่ละข้อมูลมีการกระจายความน่าจะเป็นในแต่ละกลุ่ม โดยประมาณค่าเฉลี่ยถ่วงน้ำหนักจากความน่าจะเป็นของข้อมูลในกลุ่มนั้น

Melnykov, Chen et al. (2012) ศึกษาแพคเกจ “MixSim” ขึ้นมาสำหรับโปรแกรม R ซึ่งเป็นเครื่องมือช่วยในการจำลองข้อมูลที่มีการแจกแจงปกติแบบผสม ที่มีอัตราการซ้อนทับเฉลี่ยที่ต่างกัน จากการศึกษาแพคเกจ “MixSim” จึงได้ทำการจำลองข้อมูลที่มีการแจกแจงปกติแบบผสมแล้วนำมาทำการจัดกลุ่มข้อมูลด้วยวิธีการจัดกลุ่มทั้ง 4 วิธี และเปรียบเทียบวิธีการจัดกลุ่มข้อมูลด้วยค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน ของวิธีการปรับค่าดัชนีของแรนด์ (Adjusted Rand Index), ค่าสัดส่วนของการจำแนกที่ถูกต้อง (Proportion of correct classification) และค่าการเปลี่ยนแปลงข้อมูล (Variation in Information) พบว่า วิธีการจัดกลุ่มแบบเคมีน เป็นวิธีการจัดกลุ่มที่ดีที่สุด แต่เมื่อค่าการซ้อนทับเฉลี่ยที่มีขนาดปานกลาง และขนาดใหญ่ วิธีการจัดกลุ่มแบบเคมีน ดีกว่าวิธีการจัดกลุ่มแบบเคเมดอย เพียงเล็กน้อยเท่านั้น

Hossein Parsaei ทำการรวบรวมวิธีการประมาณจำนวนกลุ่ม (Number of Clusters) ที่มีประสิทธิภาพดีที่สุด 7 วิธี ดังนี้ 1.วิธีGap statistics 2.Prediction strength method 3.Jump method 4.Calinski and Harabasz method (Pseudo F) 5.Hartigan method 6.Krzanowski and Lai method และ 7.Silhouette width ซึ่งวิธีการประมาณจำนวนกลุ่มเหล่านี้มาจากงานวิจัยของ Miligan G.W. และ Cooper M.C. (1985) ทำการศึกษาวิธีการประมาณจำนวนกลุ่มของวิธีการจัดกลุ่มข้อมูลทั้งหมด 30 วิธี สำหรับข้อมูลที่ทำการจัดกลุ่มด้วยเทคนิคของวิธีการจัดกลุ่มแบบลำดับชั้น

จากงานวิจัยของ Melnykov V. Chen WC. และ Maitra R. ผู้วิจัยนำมาศึกษาต่อ โดยทำการจัดกลุ่มข้อมูลด้วยวิธีการจัดกลุ่ม 4 วิธี คือ คือ วิธีการจัดกลุ่มแบบลำดับชั้น, วิธีการจัดกลุ่มแบบเคมีน, วิธีการจัดกลุ่มแบบฟัซซี่ซี-มีน และวิธีการจัดกลุ่มแบบอัลกอริทึม EM และจากงานวิจัยของ Miligan และ Cooper (1985) ผู้วิจัยได้นำวิธีการประมาณจำนวนกลุ่มของวิธีการจัดกลุ่มข้อมูลมาเป็นเกณฑ์การวัดประสิทธิภาพวิธีการจัดกลุ่มข้อมูล ดังนี้ 1.วิธี Calinski and Harabasz index (Pseudo F) และ 2.วิธี Silhouette width เพื่อเปรียบเทียบประสิทธิภาพวิธีการจัดกลุ่มข้อมูลที่ดีที่สุด โดยพิจารณาจากกรณีที่จำนวนกลุ่มของการจัดกลุ่มข้อมูล เท่ากับจำนวนกลุ่มของการซ้อนทับ

1.2 วัตถุประสงค์

เพื่อเปรียบเทียบประสิทธิภาพวิธีการจัดกลุ่มสำหรับข้อมูลที่มีการแจกแจงปกติแบบผสม

1.3 ข้อตกลงเบื้องต้น

การศึกษาครั้งนี้มีข้อตกลงเบื้องต้นสำหรับการดำเนินงานวิจัยดังนี้

ศึกษาตัวแบบที่มีการแจกแจงปกติหลายตัวแปร (Multivariate Normal Distribution) นั่นคือ $X=\left[x_1,x_2,...,x_p\right]\sim N\left(\mu,\Sigma\right)$ ซึ่งฟังก์ชันความน่าจะเป็นของ X อยู่ในรูปของ

$$\phi\left(\mathrm{X};\mu,\Sigma\right)=\frac{1}{\left(2\pi\right)^{p/2}\left|\Sigma\right|^{1/2}}\exp\left\{-\frac{1}{2}\left(x-\mu\right)'\Sigma^{-1}\left(x-\mu\right)\right\}$$

เมื่อ $-\infty<X<\infty;-\infty<\mu<\infty;\Sigma>0$ และ p คือ จำนวนตัวแปร

1.4 ขอบเขตของการวิจัย

ในการวิจัยครั้งนี้จะทำการศึกษาภายใต้ขอบเขตดังนี้

1. ข้อมูลมีรูปแบบการแจกแจงปกติแบบผสม (Gaussian Mixture Model)

2. ข้อมูลที่ถูกจำลองมีอัตราการซ้อนทับเฉลี่ย $(\bar{\omega})$ ที่ 1%, 10%, 20%, 30% และ 40% (Clusters with 1%, 10%, 20%, 30% and 40% Overlapping of average$(\bar{\omega})$)

3. กำหนดจำนวนกลุ่มของการซ้อนทับกัน คือ 2, 3 และ 4 กลุ่ม ตัวแปรที่ใช้ศึกษา คือ 2 และ 3 ตัวแปร ข้อมูลมีขนาดกลุ่มละ 50, 100 และ 300 ตัวอย่าง และจำนวนกลุ่มของการจัดกลุ่มข้อมูลเท่ากับ 2, 3 และ 4 กลุ่ม

4. นำวิธี Pseudo F และวิธี Silhouette width มาเป็นดัชนีการตัดสินใจ

5. ทำการจำลองข้อมูลให้มีสถานการณ์ที่แตกต่างกันตามข้อกำหนดข้างต้น โดยจำลองในแต่ละสถานการณ์จำนวน 1,000 รอบ

1.5 วิธีดำเนินการวิจัย

1. ศึกษาค้นคว้าเอกสารและข้อมูลที่เกี่ยวข้องกับงานวิจัย

2. กำหนดเงื่อนไขและขอบเขตของการวิจัย

2.1 กำหนดลักษณะการซ้อนทับกันของข้อมูล

2.2 กำหนดขนาดตัวอย่าง (n)

2.3 กำหนดจำนวนตัวแปรที่ใช้ในการศึกษา

2.4 กำหนดจำนวนกลุ่มของการจัดกลุ่มข้อมูล (k)

3. จำลองข้อมูลตามการแจกแจงและขอบเขตที่ต้องการศึกษา

3.1 ข้อมูลมีลักษณะการแจกแจงปกติแบบผสม มีจำนวนกลุ่มของการซ้อนทับกัน 2, 3 และ 4 กลุ่ม มีตัวแปรที่ใช้ในการศึกษา 2 และ 3 ตัวแปร ข้อมูลมีขนาดกลุ่มละ 50, 100 และ 300

3.2 ข้อมูลที่ถูกจำลองมีการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1%, 10%, 20%, 30% และ 40% โดยใช้ฟังก์ชัน “MixSim” ในโปรแกรม R

4. ทำการจัดกลุ่มข้อมูลด้วยวิธีการจัดกลุ่มแบบลำดับชั้น, วิธีการจัดกลุ่มแบบเคมีน, วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน และวิธีการจัดกลุ่มแบบอัลกอริทึม EM มีจำนวนกลุ่มของการจัดกลุ่มข้อมูลเท่ากับ 2, 3 และ 4 กลุ่ม

5. นำวิธี Pseudo F และวิธี Silhouette width เป็นดัชนีการตัดสินใจในแต่ละวิธีการจัดกลุ่มว่าข้อมูลควรมีจำนวนกลุ่มของการจัดกลุ่มข้อมูลเท่าไหร่

6. ทำซ้ำ 1,000 ครั้ง คำนวนหาเปอร์เซ็นต์ของวิธีการจัดกลุ่มในแต่ละวิธีจากดัชนีการตัดสินใจ

7. วัดประสิทธิภาพของวิธีการจัดกลุ่มว่าข้อมูล โดยพิจารณาจากกรณีที่จำนวนกลุ่มของการจัดกลุ่มข้อมูล เท่ากับจำนวนกลุ่มของการซ้อนทับ และสรุปผล

1.6 ประโยชน์ที่คาดว่าจะได้รับ

เพื่อใช้เป็นแนวทางสำหรับผู้วิเคราะห์ต่อการตัดสินใจเลือกใช้วิธีการจัดกลุ่มข้อมูลสำหรับข้อมูลที่มีการแจกแจงปกติแบบผสม

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

บทที่ 2

ทฤษฎีและตัวสถิติที่เกี่ยวข้อง

2.1 การแจกแจงปกติหลายตัวแปร (Multivariate Normal distribution)

คือ การแจกแจงที่มาจากการแจกแจงปกติที่มี 1 มิติไปยังการแจกแจงปกติที่มีหลายมิติ โดยที่ $X=\left[x_1,x_2,...,x_p\right]\sim N\left(\mu,\Sigma\right)$ ซึ่งมีฟังก์ชันความน่าจะเป็นของ X อยู่ในรูปของ

$$\phi\left(\mathrm{X};\mu,\Sigma\right)=\frac{1}{\left(2\pi\right)^{p/2}\left|\Sigma\right|^{1/2}}\exp\left\{-\frac{1}{2}\left(x-\mu\right)'\Sigma^{-1}\left(x-\mu\right)\right\}$$

เมื่อ $-\infty<X<\infty;-\infty<\mu<\infty;\Sigma>0$ และ p คือ จำนวนตัวแปร

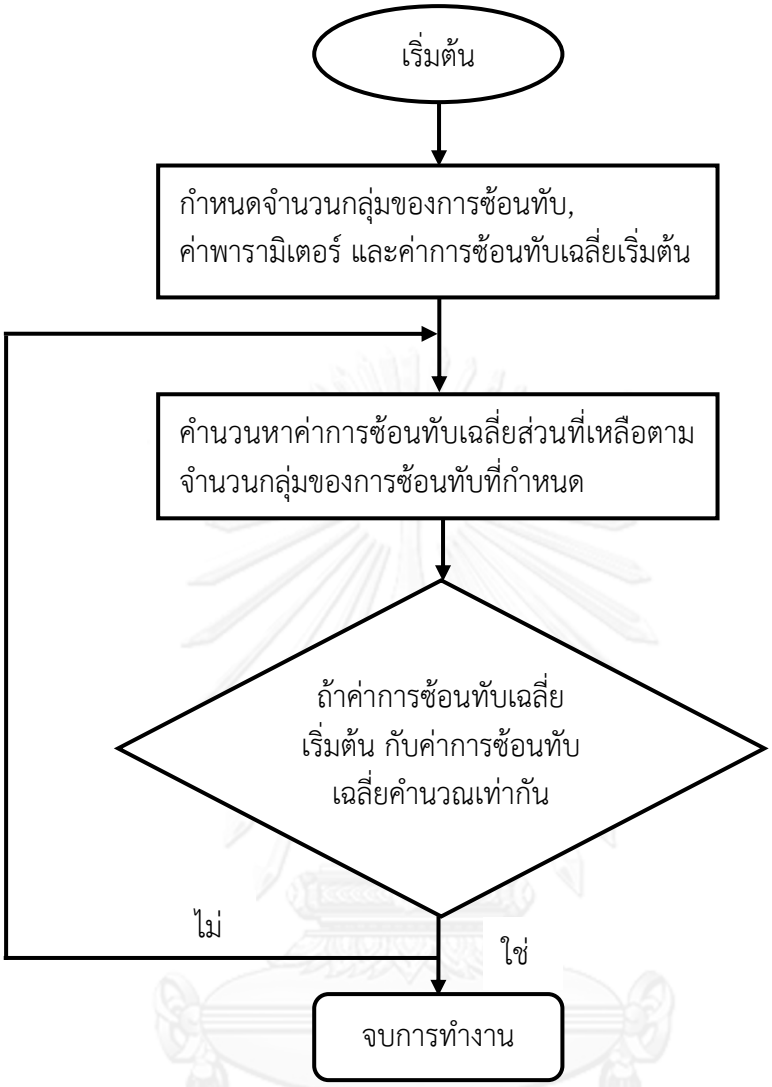
2.2 เมทริกซ์ความแปรปรวนร่วม (Covariance Matrix)

เมทริกซ์ความแปรปรวนร่วม $\left(\Sigma\right)$ สำหรับการแจกแจงปกติหลายตัวแปร สามารถแบ่งตามลักษณะของข้อมูลออกเป็น 2 กรณี ดังนี้ 1.กรณีที่ฐานข้อมูลอยู่ในรูปแบบวงกลม (Spherical or Isotopic) คือ การที่เมทริกซ์ความแปรปรวนร่วม อยู่ในรูปแบบ $\Sigma=\sigma^2I$ โดยที่ σ^2 คือ ความแปรปรวน และ I คือเมทริกซ์เอกลักษณ์ 2.กรณีที่ฐานข้อมูลอยู่ในรูปแบบวงรี (Non Spherical) คือ การที่เมทริกซ์ความแปรปรวนร่วม อยู่ในรูปแบบ $\Sigma\neq\sigma^2I$

2.3 แพคเกจ MixSim

แพคเกจ “MixSim” นำเสนอโดย Melnykov V. Chen WC. และ Maitra R. ในปี 2012 ซึ่งมีขั้นตอนในการจำลองข้อมูลดังนี้

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University



ภาพที่ 2.3.1 ขั้นตอนในการจำลองข้อมูล

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

2.4 การจัดกลุ่มแบบลำดับชั้น (Hierarchical Clustering)

การจัดกลุ่มแบบลำดับชั้น (Johnson, 1967) เป็นวิธีการจัดกลุ่มที่นิยม คำนวณง่าย นำไปใช้ในการจัดกลุ่มข้อมูลที่มีคุณลักษณะซ้อน (Nested) การจัดกลุ่มแบบลำดับขั้นสามารถแบ่งออกเป็น 2 ประเภท คือ

1. การเกาะกลุ่ม (Agglomerative) มีขั้นตอนในการทำงาน คือ เริ่มจากการจัดกลุ่มข้อมูลเป็นจำนวน n กลุ่ม จากข้อมูล n ตัว เป็นบัพภายนอก (External node) จากนั้นทำการรวมข้อมูลที่มีค่าระยะห่างระหว่าง 2 กลุ่มน้อยที่สุด โดยรวมข้อมูลครั้งละ 2 กลุ่ม ทำซ้ำจนกระทั่งสุดท้ายรวมข้อมูลเหลือเพียง 1 กลุ่ม ซึ่งเป็นบัพราก (Root node)

2. การแตกกลุ่ม (Divisive) มีขั้นตอนในการทำงานคล้ายวิธีการเกาะกลุ่ม แต่มีความแตกต่างกันที่วิธีการแตกกลุ่มนี้ จะทำงานในทิศตรงข้ามกัน คือ การแตกกลุ่มจะเริ่มจากบัพรากไปสู่บัพภายนอก หรือจากบนลงล่างนั่นเอง ซึ่งวิธีนี้ไม่เป็นที่นิยม เนื่องจากยุ่งยากและใช้เวลาในการคำนวณมากกว่าวิธีของการเกาะกลุ่ม (Lebart and Rajman, 2000)

ขั้นตอนของวิธีการจัดกลุ่มแบบลำดับชั้น โดยประเภทการเกาะกลุ่มมีขั้นตอนดังนี้

(1) แบ่งข้อมูลออกเป็น n กลุ่ม แต่ละกลุ่มมีสมาชิกเพียง 1 ตัว

(2) รวมข้อมูล 2 กลุ่มเข้าด้วยกัน โดยพิจารณาจากค่าระยะห่างระหว่างข้อมูล และเกณฑ์การรวมกลุ่มแบบ Ward Linkage

(3) พิจารณาว่าควรรวมข้อมูลกลุ่มที่ 3 เข้ากับข้อมูล 2 กลุ่มแรก หรือรวมเข้ากับข้อมูลกลุ่มอื่น โดยใช้เกณฑ์ในการรวมกลุ่มเดียวกันกับขั้นตอนที่ 2

(4) ทำขั้นตอนที่ 3 ซ้ำไปเรื่อย ๆ กระทั่งข้อมูลทุกกลุ่มรวมอยู่ในกลุ่มเดียวกัน และข้อมูลใดที่ถูกจัดกลุ่มแล้วจะไม่มีการเปลี่ยนแปลงอีก

(5) สร้างกราฟเดนโดรแกรม และกำหนดค่าระยะห่าง เพื่อทำการจัดกลุ่มข้อมูล

การใช้กราฟเดนโดรแกรม (Dendrogram) เป็นกรณีที่ใช้สำหรับเทคนิควิเคราะห์กลุ่มแบบขั้นตอนด้วยวิธีการเกาะกลุ่ม ซึ่งเป็นการลดจำนวนกลุ่มครั้งละ 1 กลุ่ม การตัดสินใจว่าควรมีกี่กลุ่มจะใช้พิจารณาจากการกำหนดค่าระยะห่างที่ตัดผ่านกราฟเดนโดรแกรม

เกณฑ์ในการรวมกลุ่มสำหรับวิธีการเกาะกลุ่มที่นำมาศึกษา คือ การรวมกลุ่มแบบ Ward Linkage เป็นการหาค่าระยะห่างของผลรวมกำลังสองภายในกลุ่ม (Sum of square within cluster distance) เพิ่มขึ้นน้อยที่สุด สามารถได้คำนวณดังนี้

$$d(v,w)=\left(\frac{n_v n_w}{n_v+n_w}\right)^{\frac{1}{2}}\left(\sum_{r=1}^{p}\left(\bar{x}_{vr}-\bar{x}_{wr}\right)^2\right)^{\frac{1}{2}}$$

โดยที่ $d(v,w)$ แทน ค่าวิธีการจัดกลุ่มแบบ Ward ของกลุ่มที่ v และกลุ่มที่ w

$\bar{x}_{vr}$ แทน ค่าเฉลี่ยของข้อมูลในกลุ่มที่ v ตัวแปรที่ r

$\bar{x}_{jr}$ แทน ค่าเฉลี่ยของข้อมูลในกลุ่มที่ w ตัวแปรที่ r

n_v แทน จำนวนข้อมูลในกลุ่มที่ v

n_w แทน จำนวนข้อมูลในกลุ่มที่ w

ข้อดีและข้อเสียของวิธีการจัดกลุ่มแบบลำดับชั้น

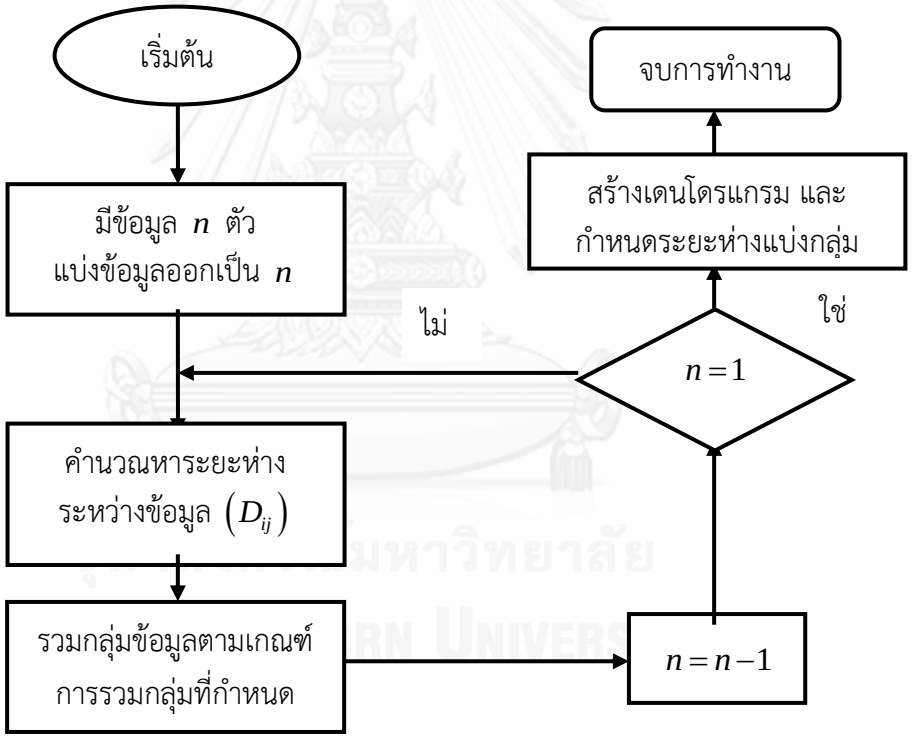
1. ข้อดี

- ไม่ต้องทราบจำนวนกลุ่มมาก่อน
- ไม่ต้องกำหนดจำนวนกลุ่มเริ่มต้น
- คำนวณได้ง่าย และใช้เวลาน้อย

2. ข้อเสีย

- มีจำนวนกลุ่มที่ไม่ชัดเจน เพราะขึ้นอยู่กับการกำหนดระยะห่างการแบ่งกลุ่ม เพื่อตัดผ่านกราฟเดนโดรแกรม
- ได้รับผลกระทบจากค่าผิดปกติ (outlier) ได้ง่าย

ขั้นตอนวิธีการจัดกลุ่มแบบลำดับขั้นประเภทการเกาะกลุ่ม



ภาพที่ 2.4.1 ขั้นตอนการจัดกลุ่มแบบลำดับชั้นประเภทการเกาะกลุ่ม

ตัวอย่างที่ 2.4.1 วิธีการจัดกลุ่มแบบลำดับขั้นประเภทการเกาะกลุ่ม โดยสมมติข้อมูล 4 รายการ คือ A, B, C และ D ในแต่ละชุดข้อมูลมีตัวแปร 2 ตัวแปร คือ r_1 และ r_2

ข้อมูล	ตัวแปร	
	r_1	r_2
A	2	4
B	4	-1
C	-1	3
D	5	-2

สามารถแสดงข้อมูลในรูปเมทริกซ์ X ขนาด 4×2 ได้ดังนี้

$$X=\begin{bmatrix} 2 & 4 \\ 4 & -1 \\ -1 & 3 \\ 5 & -2 \end{bmatrix}$$

(1) คำนวณหาค่าระยะห่างระหว่างข้อมูลแบบยุคลิด โดยใช้สมการดังนี้

$$D_{vw}=\sqrt{\sum_{r=1}^{p}\left(\vec{x}_{vr}-\vec{x}_{wr}\right)^2}$$

จากสูตรสามารถคำนวณค่าระยะห่างระหว่างข้อมูลแต่ละคู่จากเมทริกซ์ X ดังนี้

$$D(A,B)=\sqrt{(2-4)^2+(4+1)^2}=\sqrt{29}=5.3852$$
$$D(A,\mathrm{C})=\sqrt{(2+1)^2+(4-3)^2}=\sqrt{10}=3.1623$$
$$D(A,\mathrm{D})=\sqrt{(2-5)^2+(4+2)^2}=\sqrt{45}=6.7082$$
$$D(B,C)=\sqrt{(4+1)^2+(-1-3)^2}=\sqrt{41}=6.4031$$
$$D(B,D)=\sqrt{(4-5)^2+(-1+2)^2}=\sqrt{2}=1.4142$$
$$D(C,D)=\sqrt{(-1-5)^2+(3+2)^2}=\sqrt{61}=7.8102$$

ค่า D_{vw} ทั้งหมดที่คำนวณได้ สามารถแสดงในรูปเมทริกซ์ระยะห่าง $\left(\tilde{D}\right)$ ดังนี้

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

$$\tilde{D} = \begin{array}{c} \\ A \\ B \\ C \\ D \end{array} \begin{array}{c} \begin{array}{cccc} A & B & C & D \end{array} \\ \begin{bmatrix} 0 & 5.3852 & 3.1623 & 6.7082 \\ 5.3852 & 0 & 6.4031 & 1.4142 \\ 3.1623 & 6.4031 & 0 & 7.8102 \\ 6.7082 & 1.4142 & 7.8102 & 0 \end{bmatrix} \end{array}$$

(2) ทำการแบ่งกลุ่มด้วยวิธีการจัดกลุ่มแบบ Ward Linkage โดยใช้สมการดังนี้

$$d(v,w) = \left(\frac{n_v n_w}{n_v + n_w} \right)^{\frac{1}{2}} \left(\sum_{r=1}^{p} \left(\bar{x}_{vr} - \bar{x}_{wr} \right) \right)^{\frac{1}{2}}$$

ครั้งที่ 1 จากค่าของเมทริกซ์ $\tilde{D}$ ซึ่งเป็นค่าระยะห่างระหว่างข้อมูลแต่ละหน่วย นำมาคำนวณหาค่าวิธีการจัดกลุ่มแบบ Ward Linkage ซึ่งสามารถคำนวณได้ดังนี้

$$d(\mathrm{A},\mathrm{B}) = \sqrt{\frac{(1)(1)}{1+1}}(5.3852) = 3.8079$$

$$d(\mathrm{A},\mathrm{C}) = \sqrt{\frac{(1)(1)}{1+1}}(3.1623) = 2.2361$$

$$d(\mathrm{A},\mathrm{D}) = \sqrt{\frac{(1)(1)}{1+1}}(6.7082) = 4.7434$$

$$d(\mathrm{B},\mathrm{C}) = \sqrt{\frac{(1)(1)}{1+1}}(6.4031) = 4.5277$$

$$d(B,D) = \sqrt{\frac{(1)(1)}{1+1}}(1.4142) = 1.0000$$

$$d(\mathrm{C},D) = \sqrt{\frac{(1)(1)}{1+1}}(7.8102) = 5.5226$$

จะเห็นว่า $d(B,D)$ มีค่าน้อยที่สุด ดังนั้นจึงเลือกข้อมูล B และข้อมูล D เข้าไปอยู่ในกลุ่มเดียวกัน ที่มีค่าระยะห่างเท่ากับ 1.0000

ครั้งที่ 2 จากการรวมกลุ่มข้อมูลในครั้งที่ 1 ทำให้เกิดเมทริกซ์ของข้อมูล X ใหม่ซึ่งเกิดจากการรวมกลุ่มของข้อมูล โดยนำค่าของข้อมูลที่ทำการรวมกลุ่มมาทำการหาตัวแทนกลุ่มใหม่

$$X = \begin{bmatrix} 2 & 4 \\ 4.5 & -1.5 \\ -1 & 3 \end{bmatrix}$$

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

สามารถคำนวณได้ดังนี้

ค่าเฉลี่ยข้อมูล B และข้อมูล D ในตัวแปร $r_1 = \frac{(4+5)}{2} = 4.5$

ค่าเฉลี่ยข้อมูล B และข้อมูล D ในตัวแปร $r_2 = \frac{(-1-2)}{2} = -1.5$

ดังนั้นสามารถคำนวณหาค่าระยะห่างระหว่างกลุ่มใหม่ ได้ดังนี้

$$D\left((BD),\mathrm{A}\right) = \sqrt{(4.5-2)^2 + (-1.5-4)^2} = \sqrt{36.5} = 6.0415$$

$$D\left((BD),\mathrm{C}\right) = \sqrt{(4.5+1)^2 + (-1.5-3)^2} = \sqrt{50.5} = 7.1063$$

$$D\left(A,\mathrm{C}\right) = \sqrt{(2+1)^2 + (4-3)^2} = \sqrt{10} = 3.1623$$

ค่า D_{vw} ทั้งหมดที่คำนวณได้ สามารถแสดงในรูปเมทริกซ์ระยะห่าง $\left(\tilde{D}\right)$ ดังนี้

$$\tilde{D} = \begin{array}{c} \\ A \\ (BD) \\ C \end{array} \begin{array}{c} \begin{array}{ccc} A & (BD) & C \end{array} \\ \begin{bmatrix} 0 & 6.0415 & 3.1623 \\ 6.0415 & 0 & 7.1063 \\ 3.1623 & 7.1063 & 0 \end{bmatrix} \end{array}$$

นำค่าระยะห่างระหว่างกลุ่มของข้อมูลใหม่ มาคำนวณหาค่าวิธีการจัดกลุ่มแบบ Ward Linkage ซึ่งสามารถคำนวณได้ดังนี้

$$d\left((BD),A\right) = \sqrt{\frac{(2)(1)}{(2+1)}}\left(6.0415\right) = 4.9329$$

$$d\left((BD),\mathrm{C}\right) = \sqrt{\frac{(2)(1)}{(2+1)}}\left(7.1063\right) = 5.8023$$

$$d\left(\mathrm{A},\mathrm{C}\right) = \sqrt{\frac{(1)(1)}{(1+1)}}\left(3.1623\right) = 2.2361$$

จะเห็นว่า $d\left(\mathrm{A},\mathrm{C}\right)$ มีค่าน้อยที่สุด ดังนั้นจึงเลือกข้อมูล A และข้อมูล C เข้าไปอยู่ในกลุ่มเดียวกัน ที่มีค่าระยะห่างเท่ากับ 2.2361

ครั้งที่ 3 จากการรวมกลุ่มข้อมูลในครั้งที่ 2 ทำให้เกิดเมทริกซ์ของข้อมูล X ใหม่ซึ่งเกิดจากการรวมกลุ่มของข้อมูล โดยนำค่าของข้อมูลที่ทำการรวมกลุ่มมาทำการเฉลี่ย

$$X = \begin{bmatrix} 0.5 & 3.5 \\ 4.5 & -1.5 \end{bmatrix}$$

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

สามารถคำนวณได้ดังนี้

ค่าเฉลี่ยข้อมูล A และข้อมูล C ในตัวแปร $r_1 = \frac{(2-1)}{2} = 0.5$

ค่าเฉลี่ยข้อมูล A และข้อมูล C ในตัวแปร $r_2 = \frac{(4+3)}{2} = 3.5$

ดังนั้นสามารถคำนวณหาค่าระยะห่างระหว่างกลุ่มใหม่ ได้ดังนี้

$$D((BD),(AC)) = \sqrt{(0.5-4.5)^2 + (3.5+1.5)^2} = \sqrt{41} = 6.4031$$

ค่า D_{ij} ทั้งหมดที่คำนวณได้ สามารถแสดงในรูปเมทริกซ์ระยะห่าง $(\tilde{D})$ ดังนี้

$$\tilde{D} = \begin{matrix} \\ (AC) \\ (BD) \end{matrix} \begin{matrix} \begin{matrix} (AC) & (BD) \end{matrix} \\ \begin{bmatrix} 0 & 6.4031 \\ 6.4031 & 0 \end{bmatrix} \end{matrix}$$

นำค่าระยะห่างระหว่างกลุ่มของข้อมูลใหม่ มาคำนวณหาค่าวิธีการจัดกลุ่มแบบ Ward Linkage ซึ่งสามารถคำนวณได้ดังนี้

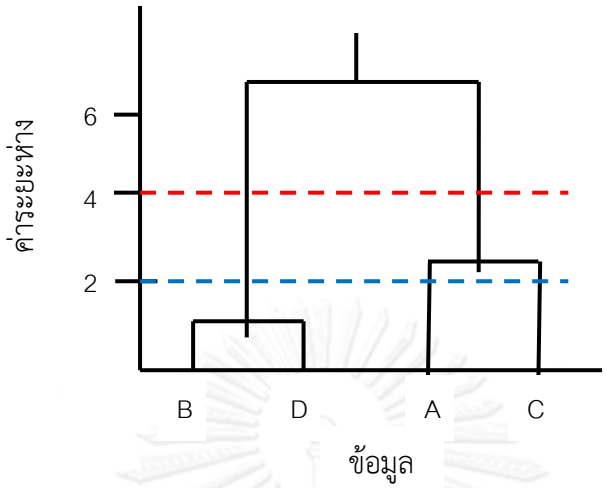
$$d((AC),(BD)) = \sqrt{\frac{(2)(2)}{(2+2)}}(6.4031) = 6.4031$$

จะเห็นว่าในครั้งที่ 3 เมื่อนำกลุ่ม (AC) และกลุ่ม (BD) มาทำการรวมกลุ่มกันแล้ว พบว่าข้อมูลทั้งหมดถูกรวมเข้ามาอยู่ในกลุ่มเดียวกัน จึงนำค่าระยะห่างที่คำนวณจากวิธีการจัดกลุ่มแบบWard Linkage มาสร้างเดนโดรแกรม และหาระยะในการแบ่งกลุ่มมาทำการแบ่งกลุ่ม โดยสามารถสรุปเป็นตาราง ดังนี้

ตารางแสดงค่าระยะห่างที่คำนวณจากวิธีการจัดกลุ่มแบบ Ward Linkage

ครั้งที่	กลุ่มที่ 1	กลุ่มที่ 2	ระยะห่าง
1	B	D	1.0000
2	A	C	2.2361
3	(AC)	(BD)	6.4031

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University



ภาพที่ 2.4.2 กราฟเดนโดรแกรม (Dendrogram)

(3) จากกราฟเดนโดรแกรม เส้นตัดแกนตั้ง คือ ค่าระยะห่างที่คำนวณจากวิธีการจัดกลุ่มแบบ Ward Linkage ส่วนเส้นตัดแกนนอน คือ ข้อมูลที่จะทำการจัดกลุ่ม จากกราฟ ถ้ากำหนดค่าระยะห่างในการแบ่งกลุ่ม 4.0000 (เส้นตัดสีแดง) จะสามารถแบ่งข้อมูลได้ 2 กลุ่ม คือ กลุ่ม (AC) และ กลุ่ม (BD) แต่ถ้ากำหนดค่าระยะห่างในการแบ่งกลุ่ม 2.0000 (เส้นตัดสีน้ำเงิน) จะสามารถแบ่งข้อมูลได้ 3 กลุ่ม คือ กลุ่ม A, กลุ่ม (BD) และ กลุ่ม C

2.5 วิธีการจัดกลุ่มแบบเคมีน (K-Means Clustering)

เทคนิคการจัดกลุ่มข้อมูลโดยอัลกอริทึมการจัดกลุ่มแบบเคมีน เป็นวิธีการจัดกลุ่มแบบแบ่งส่วน (Partitioning) ซึ่งสมาชิกแต่ละตัวภายในกลุ่มจะมีระยะอยู่ใกล้จุดศูนย์กลาง หรือตัวแทนกลุ่ม (Centroid) มากที่สุด สามารถสรุปขั้นตอนได้ดังนี้

(1) กำหนดจำนวนกลุ่มที่ต้องการจะแบ่ง และสุ่มจุดศูนย์กลางหรือตัวแทนของแต่ละกลุ่ม

(2) คำนวณหาค่าระยะห่างของข้อมูลทุกตัวกับตัวแทนกลุ่มในแต่ละกลุ่ม และพิจารณาค่าระยะห่างที่น้อยที่สุด

(3) ปรับจุดศูนย์กลางของข้อมูลแต่ละกลุ่มใหม่ โดยคำนวณจากค่าเฉลี่ยของข้อมูลภายในกลุ่ม $(\bar{x}_z)$ เพื่อใช้เป็นตัวแทนใหม่ของกลุ่ม

$$\bar{x}_z = \frac{1}{n_z}\sum_{i=1}^{n_z} x_{iz}; i = 1,2,...,\mathrm{n}_z; \mathrm{z} = 1,2,...,\mathrm{k}$$

โดยที่ n_z คือ จำนวนข้อมูลในกลุ่มแต่ละครั้งของการจัดกลุ่ม

x_{iz} คือ ข้อมูลในหน่วยที่ i กลุ่มที่ z

$\bar{x}_z$ คือ ค่าเฉลี่ยของข้อมูลในกลุ่มที่ z

k คือ จำนวนกลุ่มทั้งหมด

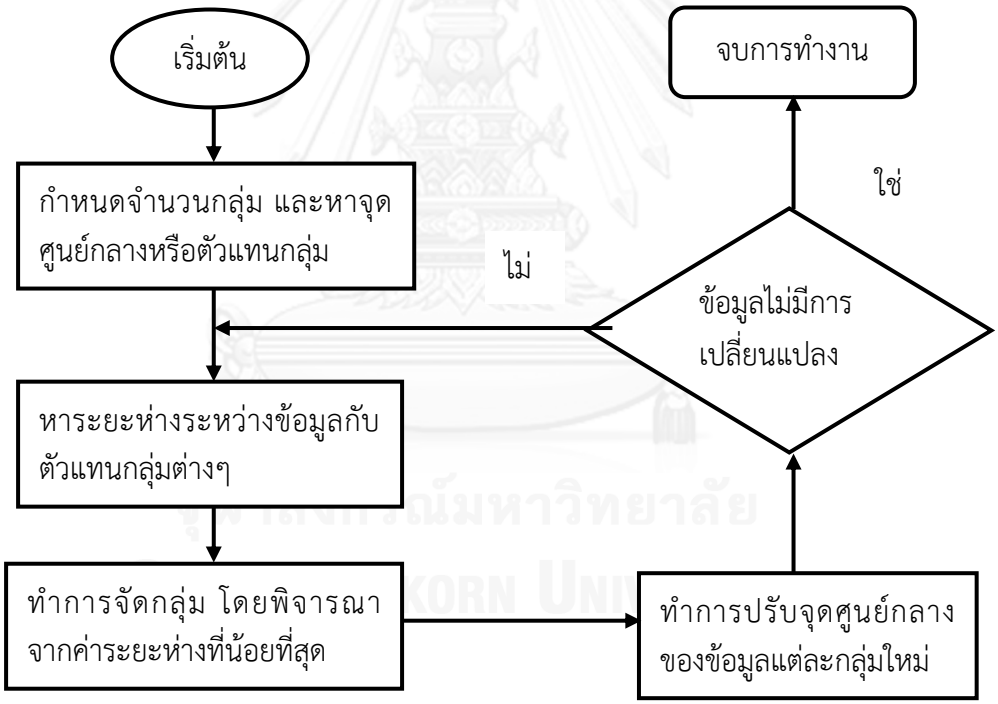
(4) ทำขั้นตอนที่ (2) และ (3) ซ้ำ กระทั่งสมาชิกทุกตัวภายในกลุ่มแต่ละกลุ่มไม่มีการเปลี่ยนแปลงไปยังกลุ่มอื่น ๆ

ข้อดีและข้อเสียของวิธีการจัดกลุ่มแบบเคมีน

1. ข้อดี
   - ใช้งานได้ดีกับข้อมูลที่เป็น Spherical

2. ข้อเสีย
   - การกำหนดจำนวนกลุ่ม และตัวแทนกลุ่มเริ่มต้น มีผลต่อประสิทธิภาพการจัดกลุ่ม
   - ใช้เวลาในการคำนวณมาก

ขั้นตอนวิธีการจัดกลุ่มบบเคมีน



ภาพที่ 2.5.1 ขั้นตอนการจัดกลุ่มแบบเคมีน

ตัวอย่างที่ 2.5.1 วิธีการจัดกลุ่มแบบเคมีน โดยสมมติข้อมูลมี 4 รายการ คือ A, B, C และ D ในแต่ละชุดข้อมูลมีตัวแปร 2 ตัวแปร คือ r_1 และ r_2 กำหนดจำนวนกลุ่มที่ต้องการแบ่งเท่ากับ 2 กลุ่ม

ข้อมูล	ตัวแปร	
	r_1	r_2
A	2	4
B	4	-1
C	-1	3
D	5	-2

(1) เนื่องจากกำหนดจำนวนกลุ่มที่ต้องการแบ่งเท่ากับ 2 กลุ่ม จึงทำการสุ่มเลือกตัวแทนกลุ่มมา 2 จุด โดยจัดให้ข้อมูล A กับข้อมูล B อยู่กลุ่มเดียวกัน และข้อมูล C กับข้อมูล D อยู่ในกลุ่มเดียวกัน ดังนี้

กลุ่ม	ตัวแปร	
	r_1	r_2
1	$\frac{2+4}{2}=3$	$\frac{4+(-1)}{2}=1.5$
2	$\frac{(-1)+5}{2}=2$	$\frac{3+(-2)}{2}=0.5$

ในที่นี้กำหนดให้ R_z คือ ตัวแทนของกลุ่มข้อมูล z

(2) คำนวณหาค่าระยะห่างระหว่างข้อมูลแบบยุคลิด กับตัวแทนของกลุ่ม (D_{iz}) ที่สุ่มเลือกมาจากขั้นตอนที่ 1 ในที่นี้มี 2 ตัวแปร $(p=2)$

$$D_{iz}=\sqrt{\sum_{r=1}^{2}\left(x_{ir}-R_{zr}\right)^2}$$

โดยค่า D_{ij} คือ ค่าระยะห่างระหว่างข้อมูลกับตัวแทนกลุ่ม เมื่อพิจารณาค่าระยะห่างที่จุดข้อมูล A คือ $D(A,R_1)$ และ $D(A,R_2)$ พิจารณาข้อมูลไปอยู่ในกลุ่มเดียวกับตัวแทนกลุ่มจากค่าระยะห่างที่น้อยที่สุด

$$D(A,R_1)=\sqrt{(2-3)^2+(4-1.5)^2}=2.6926$$
$$D(A,R_2)=\sqrt{(2-2)^2+(4-0.5)^2}=3.5$$

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

จากค่า $D(A,R_1)$ และ $D(A,R_2)$ แสดงให้เห็นว่า ค่าระยะห่างระหว่างจุดข้อมูล A กับตัวแทนกลุ่ม R_1 มีค่าน้อยกว่าระยะห่างระหว่างจุดข้อมูล A กับตัวแทนกลุ่ม R_2 ดังนั้น ข้อมูล A จึงควรอยู่กลุ่ม R_1

(3) คำนวณหาค่าระยะห่างของตัวแทนกลุ่ม กับข้อมูลที่เหลือจนครบ

$$D(B,R_1)=\sqrt{(4-3)^2+(-1-1.5)^2}=2.6926$$

$$D(B,R_2)=\sqrt{(4-2)^2+(-1-0.5)^2}=2.5$$

$$D(C,R_1)=\sqrt{(-1-3)^2+(3-1.5)^2}=4.2720$$

$$D(C,R_2)=\sqrt{(-1-2)^2+(3-0.5)^2}=3.9051$$

$$D(D,R_1)=\sqrt{(5-3)^2+(-2-1.5)^2}=4.0311$$

$$D(D,R_2)=\sqrt{(5-2)^2+(-2-0.5)^2}=3.9051$$

จากค่า $D(B,R_1)$ และ $D(B,R_2)$ แสดงให้เห็นว่า ค่าระยะห่างระหว่างจุดข้อมูล B กับตัวแทนกลุ่ม R_1 มีค่ามากกว่าระยะห่างระหว่างจุดข้อมูล B กับตัวแทนกลุ่ม R_2 ดังนั้น ข้อมูล B จึงควรอยู่กลุ่ม R_2 เมื่อพิจารณาค่า $D(C,R_1)$ กับ $D(C,R_2)$ และ ค่า $D(D,R_1)$ กับ $D(D,R_2)$ ซึ่งพบว่าค่าระยะห่างระหว่างจุดข้อมูล C และจุดข้อมูล D จึงควรอยู่กลุ่ม R_2 เหมือนกัน

(4) ทำการปรับเปลี่ยนจุดศูนย์กลางของแต่ละกลุ่มใหม่ โดยหาค่าเฉลี่ยของสมาชิกภายในกลุ่ม ดังนี้

กลุ่ม	ตัวแทนกลุ่ม	
	R_1	R_1
A (BCD)	2 $\frac{4+(-1)+5}{3}=2.6667$	4 $\frac{(-1)+3+(-2)}{3}=0$

(5) นำตัวแทนกลุ่มที่ได้มาทำขั้นตอนที่ 2 ถึงขั้นตอนที่ 4 ซ้ำ กระทั่งข้อมูลไม่มีการเปลี่ยนแปลง ซึ่งมีผลดังนี้

กลุ่ม	ระยะห่างของข้อมูลภายในกลุ่ม			
	A	B	C	D
1	6.0415	2.5495	7.1063	4.5277
2	1.5811	5.7009	1.5811	7.1063

จากตาราง พบว่าข้อมูลไม่มีการเปลี่ยนแปลงไปยังกลุ่มอื่นแล้ว จึงหยุดดำเนินการ สามารถสรุปผลได้ว่า กลุ่มข้อมูลออกที่แบ่งออกเป็น 2 กลุ่ม ดังนี้ กลุ่มที่ 1 มีสมาชิก คือ ข้อมูล B กับ ข้อมูล D และกลุ่มที่ 2 มีสมาชิก คือ ข้อมูล A กับ ข้อมูล C

2.6 วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน (Fuzzy C-Means Clustering)

วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน พัฒนามาจากการจัดกลุ่มข้อมูลเป็นกลุ่มย่อย ๆ ที่ในแต่ละกลุ่มถูกแบ่งกันอย่างชัดเจน มาเป็นกลุ่มที่มีข้อมูลบางส่วนอยู่ร่วมกัน และมีความน่าจะเป็นที่จะถูกจัดเข้ากลุ่มมากกว่าหนึ่งกลุ่ม ในการทำงานจะกำหนดค่าความเป็นสมาชิกของข้อมูลแต่ละตัวกับตัวแทนของกลุ่มแต่ละกลุ่ม โดยข้อมูลถูกจัดเข้ากลุ่มที่มีค่าความเป็นสมาชิกกับตัวแทนกลุ่มสูงสุด สามารถสรุปขั้นตอนได้ดังนี้

(1) กำหนดจำนวนกลุ่มที่ต้องการแบ่งเป็น k กลุ่ม กำหนดตัวแทนกลุ่มเริ่มต้นจากการสุ่ม และกำหนดค่าระดับความเป็นสมาชิก (m) ในกลุ่ม

(2) คำนวณค่าความเป็นสมาชิก U ของข้อมูลทุกตัว โดยใช้สมการ

$$U_{ie} = \frac{1}{\sum_{z=1}^{k}\left(\frac{D_{ie}}{D_{iz}}\right)^{\frac{2}{(m-1)}}}, \quad D_{ic} = \left\| x_i - c_e \right\|$$

เมื่อ D_{ie} คือ ระยะห่างระหว่างข้อมูลตัวที่ i กับตัวแทนของกลุ่มที่ e

โดยที่ c_e คือ ตัวแทนข้อมูลกลุ่มที่ e

x_i คือ ข้อมูลตัวที่ i

m คือ ค่าระดับความเป็นสมาชิก (Fuzziness index) โดยที่ $m \in (1,\infty)$

(3) คำนวณค่าดัชนีความผิดพลาด (Cost function) และทำการแบ่งข้อมูลเข้ากลุ่ม

$$J(U, c_1, c_2, \ldots c_e) = \sum_{e=1}^{k} J_e = \sum_{e=1}^{k}\sum_{i=1}^{n} {U_{ie}}^{m} \left\| x_i - c_e \right\|^2$$

โดยที่ c_e คือ ตัวแทนกลุ่มที่ e

x_i คือ ข้อมูลตัวที่ i

m คือ ค่าความเป็นสมาชิก (Fuzziness index) โดยที่ $m \in (1,\infty)$

U_{ie} คือ ค่าระดับความเป็นสมาชิกของข้อมูลตัวที่ i ของกลุ่มที่ e

(4) คำนวณหาตัวแทนกลุ่มใหม่ และปรับค่าความเป็นสมาชิกในกลุ่ม

$$c_e = \frac{\sum_{i=1}^{n} U_{ie}^{m} x_i}{\sum_{i=1}^{n} U_{ie}^{m}}$$

โดยที่ c_e คือ ตัวแทนกลุ่มที่ e

x_i คือ ข้อมูลตัวที่ i

m คือ ค่าระดับความเป็นสมาชิก (Fuzziness index) โดยที่ $m \in (1, \infty)$

U_{ie} คือ ค่าความเป็นสมาชิกของข้อมูลตัวที่ i ของกลุ่มที่ e

(5) ทำซ้ำขั้นตอนที่ (2), (3) และ (4) กระทั่งตัวแทนกลุ่มปรับค่าเข้าสู่จุดศูนย์กลางกลุ่ม และค่าดัชนีความผิดพลาดของกลุ่มมีค่าเป็นที่น่าพอใจ

ข้อดีและข้อเสียของวิธีการจัดกลุ่มแบบฟัซซี่ซีมีน

1. ข้อดี

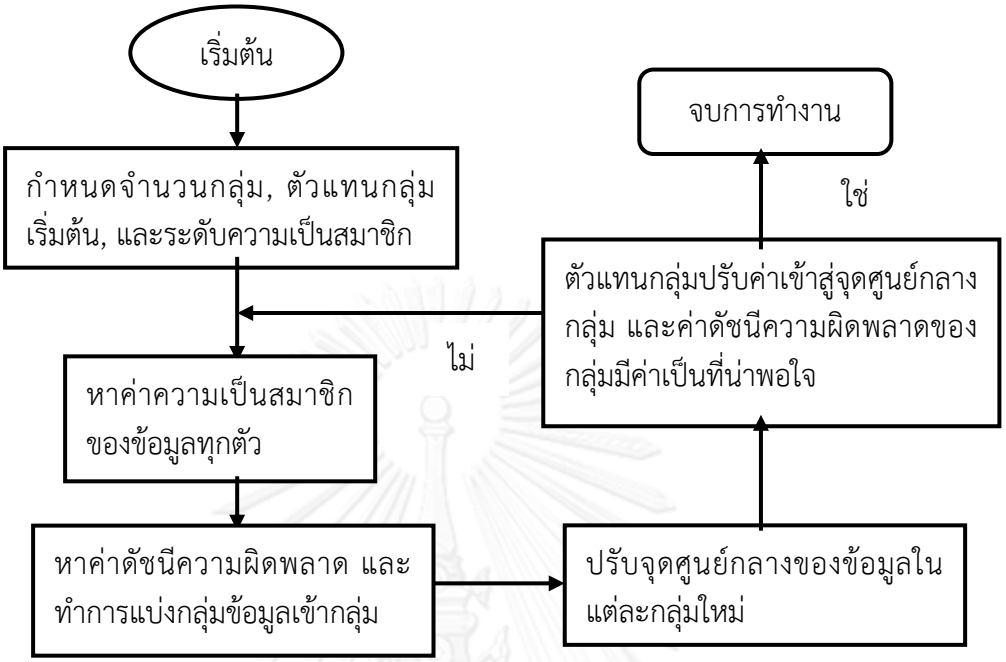
- การจัดกลุ่มของข้อมูลมีความน่าเชื่อถือ เพราะเป็นการจัดข้อมูลเข้ากลุ่มตามค่าระดับความเป็นสมาชิกในแต่ละกลุ่ม
- สามารถจำแนกกลุ่มข้อมูลที่มีความคล้ายกันมากได้

2. ข้อเสีย

- การกำหนดจำนวนกลุ่ม และค่าเริ่มต้น มีผลต่อประสิทธิภาพการจัดกลุ่ม
- ใช้เวลาในการคำนวณมาก

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ขั้นตอนการจัดกลุ่มแบบฟัซซี่ซีมีน



ภาพที่ 2.6.1 ขั้นตอนการจัดกลุ่มแบบฟัซซี่ซีมีน

ตัวอย่างที่ 2.6.1 วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน โดยสมมติข้อมูลมี 4 รายการ คือ A, B, C และ D ในแต่ละชุดข้อมูลมีตัวแปร 2 ตัวแปร คือ r_1 และ r_2 โดยกำหนดจำนวนกลุ่มที่ต้องการจะแบ่งเท่ากับ 2 กลุ่ม

ข้อมูล	ตัวแปร	
	r_1	r_2
A	2	4
B	4	-1
C	-1	3
D	5	-2

(1) ทำการแบ่งกลุ่มข้อมูลออกเป็น 2 กลุ่ม จึงสุ่มตัวแทนกลุ่มข้อมูลแต่ละกลุ่มออกมา ได้ดังนี้

กลุ่ม	ตัวแปร	
	r_1	r_2
1	-1.2	2.1
2	1.8	1.3

ในที่นี้กำหนดให้ c_1 คือ ตัวแทนของกลุ่ม 1 และ c_2 คือ ตัวแทนของกลุ่ม 2

(2) กำหนดค่าความเป็นสมาชิก (Fuzziness index) ที่ $m=1.5$ และคำนวณค่าความเป็นสมาชิก U ของข้อมูลทุกตัว กับตัวแทนกลุ่ม ที่สุ่มในขั้นตอนที่ 1 ในที่นี้มีค่าความเป็นสมาชิกเท่ากับ $m=1.5$ และจำนวนกลุ่มที่ต้องการจะแบ่งเท่ากับ 2 กลุ่ม สมการที่ใช้จึงเป็นดังนี้

$$U_{ie}=\frac{1}{\sum_{z=1}^{k}\left(\frac{D_{ie}}{D_{iz}}\right)^{\frac{2}{(m-1)}}}=\frac{1}{\sum_{z=1}^{2}\left(\frac{D_{ie}}{D_{iz}}\right)^{4}}$$

คำนวณความเป็นสมาชิกของข้อมูลแต่ละตัวกับตัวแทนกลุ่มที่สุ่มได้ จะเลือกให้ข้อมูลที่มีค่าความเป็นสมาชิกกับตัวแทนกลุ่มที่มากที่สุดไปอยู่ในกลุ่มเดียวกับตัวแทนกลุ่มนั้น ๆ

กำหนดให้ D_{1A} คือ ระยะห่างระหว่างข้อมูล A กับตัวแทนกลุ่มที่ 1 (c_1)

D_{2A} คือ ระยะห่างระหว่างข้อมูล A กับตัวแทนกลุ่มที่ 2 (c_2)

$$D_{1A}=\sqrt{\left((-1.2)-2\right)^2+\left(2.1-4\right)^2}=\sqrt{13.85}$$

$$D_{2A}=\sqrt{\left(1.8-2\right)^2+\left(1.3-4\right)^2}=\sqrt{7.33}$$

$$U_{1A}=\frac{1}{\left(\frac{D_{1A}}{D_{1A}}\right)^4+\left(\frac{D_{1A}}{D_{2A}}\right)^4}=\frac{1}{\left(\frac{\sqrt{13.85}}{\sqrt{13.85}}\right)^4+\left(\frac{\sqrt{13.85}}{\sqrt{7.33}}\right)^4}=0.2188$$

$$U_{2A}=\frac{1}{\left(\frac{D_{2A}}{D_{1A}}\right)^4+\left(\frac{D_{2A}}{D_{2A}}\right)^4}=\frac{1}{\left(\frac{\sqrt{7.33}}{\sqrt{13.85}}\right)^4+\left(\frac{\sqrt{7.33}}{\sqrt{7.33}}\right)^4}=0.7812$$

จากค่าข้างต้นเห็นว่าค่าความเป็นสมาชิกระหว่างข้อมูล A กับตัวแทนกลุ่มที่ 1 (U_{1A}) มีค่าน้อยกว่าค่าความเป็นสมาชิกระหว่างข้อมูล A กับตัวแทนกลุ่มที่ 2 (U_{2A}) ดังนั้น ข้อมูล A ควรอยู่กลุ่มเดียวกับตัวแทนกลุ่มที่ 2 (U_{2A})

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

(3) คำนวณหาค่าความเป็นสมาชิกระหว่างข้อมูล A กับตัวแทนกลุ่มที่เหลือจนครบ

กำหนดให้ D_{1B} คือ ระยะห่างระหว่างข้อมูล B กับตัวแทนกลุ่มที่ 1 (c_1)

D_{2B} คือ ระยะห่างระหว่างข้อมูล B กับตัวแทนกลุ่มที่ 2 (c_2)

$$D_{1B}=\sqrt{\left((-1.2)-4\right)^2+\left(2.1-(-1)\right)^2}=\sqrt{36.65}$$

$$D_{2B}=\sqrt{\left(1.8-4\right)^2+\left(1.3-(-1)\right)^2}=\sqrt{10.13}$$

$$U_{1B} = \frac{1}{\left(\frac{D_{1B}}{D_{1B}}\right)^4 + \left(\frac{D_{1B}}{D_{2B}}\right)^4} = \frac{1}{\left(\frac{\sqrt{36.65}}{\sqrt{36.65}}\right)^4 + \left(\frac{\sqrt{36.65}}{\sqrt{10.13}}\right)^4} = 0.8869$$

$$U_{2B} = \frac{1}{\left(\frac{D_{2B}}{D_{1B}}\right)^4 + \left(\frac{D_{2B}}{D_{2B}}\right)^4} = \frac{1}{\left(\frac{\sqrt{10.13}}{\sqrt{36.65}}\right)^4 + \left(\frac{\sqrt{10.13}}{\sqrt{10.13}}\right)^4} = 0.9290$$

กำหนดให้ D_{1C} คือ ระยะห่างระหว่างข้อมูล C กับตัวแทนกลุ่มที่ 1 (C_1)

D_{2C} คือ ระยะห่างระหว่างข้อมูล C กับตัวแทนกลุ่มที่ 2 (C_2)

$$D_{1C} = \sqrt{\left((-1.2)-(-1)\right)^2 + (2.1-3)^2} = \sqrt{0.85}$$

$$D_{2C} = \sqrt{\left(1.8-(-1)\right)^2 + (1.3-3)^2} = \sqrt{10.73}$$

$$U_{1C} = \frac{1}{\left(\frac{D_{1C}}{D_{1C}}\right)^4 + \left(\frac{D_{1C}}{D_{2C}}\right)^4} = \frac{1}{\left(\frac{\sqrt{0.85}}{\sqrt{0.85}}\right)^4 + \left(\frac{\sqrt{0.85}}{\sqrt{10.73}}\right)^4} = 0.9938$$

$$U_{2C} = \frac{1}{\left(\frac{D_{2C}}{D_{1C}}\right)^4 + \left(\frac{D_{2C}}{D_{2C}}\right)^4} = \frac{1}{\left(\frac{\sqrt{10.73}}{\sqrt{0.85}}\right)^4 + \left(\frac{\sqrt{10.73}}{\sqrt{10.73}}\right)^4} = 0.0062$$

กำหนดให้ D_{1D} คือ ระยะห่างระหว่างข้อมูล D กับตัวแทนกลุ่มที่ 1 (C_1)

D_{2D} คือ ระยะห่างระหว่างข้อมูล D กับตัวแทนกลุ่มที่ 2 (C_2)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

$$D_{1D} = \sqrt{\left((-1.2)-5\right)^2 + \left(2.1-(-2)\right)^2} = \sqrt{55.25}$$

$$D_{2D} = \sqrt{\left(1.8-5\right)^2 + \left(1.3-(-2)\right)^2} = \sqrt{21.13}$$

$$U_{1D} = \frac{1}{\left(\frac{D_{1D}}{D_{1D}}\right)^4 + \left(\frac{D_{1D}}{D_{2D}}\right)^4} = \frac{1}{\left(\frac{\sqrt{55.25}}{\sqrt{55.25}}\right)^4 + \left(\frac{\sqrt{55.25}}{\sqrt{21.13}}\right)^4} = 0.1276$$

$$U_{2D} = \frac{1}{\left(\frac{D_{2D}}{D_{1D}}\right)^4 + \left(\frac{D_{2D}}{D_{2D}}\right)^4} = \frac{1}{\left(\frac{\sqrt{21.13}}{\sqrt{55.25}}\right)^4 + \left(\frac{\sqrt{21.13}}{\sqrt{21.13}}\right)^4} = 0.8724$$

จากค่าความเป็นสมาชิกระหว่างข้อมูลที่ได้คำนวณข้างต้น พบว่า ค่าความเป็นสมาชิกระหว่างข้อมูล B กับตัวแทนกลุ่มที่ 2 มีค่ามากกว่าค่าความเป็นสมาชิกระหว่างข้อมูล B กับตัวแทนกลุ่มที่ 1 ดังนั้น ข้อมูล B จึงควรอยู่ในกลุ่มเดียวกับตัวแทนกลุ่มที่ 2 ทำนองเดียวกัน เมื่อพิจารณาค่าความเป็นสมาชิกของข้อมูล D กับตัวแทนทั้ง 2 กลุ่ม พบว่าข้อมูล D ควรอยู่ในกลุ่มเดียวกับตัวแทนกลุ่มที่ 2 เช่นกัน นั่นคือ ข้อมูล A, B และ D อยู่ในกลุ่มเดียวกับตัวแทนกลุ่มที่ 2 ในขณะที่ ค่าความเป็นสมาชิกระหว่างข้อมูล C กับตัวแทนกลุ่มที่ 1 มีค่ามากกว่าค่าความเป็นสมาชิกระหว่างข้อมูล C กับตัวแทนกลุ่มที่ 2 ดังนั้น C จึงควรอยู่ในกลุ่มเดียวกับตัวแทนกลุ่มที่ 1 นั่นคือ ข้อมูล C อยู่กลุ่มเดียวกับตัวแทนกลุ่มที่ 1

(4) เมื่อทำการจัดกลุ่มแล้ว จึงปรับเปลี่ยนตัวแทนกลุ่มแต่ละกลุ่ม โดยใช้สูตรดังต่อไปนี้

$$c_i = \frac{\sum_{i=1}^{n} U_{ie}^{m} x_i}{\sum_{i=1}^{n} U_{ie}^{m}}$$

จากที่กำหนดให้ $m = 1.5$ และ $n = 4$ สามารถปรับสูตรได้ดังนี้

$$c_i = \frac{\sum_{i=1}^{4} U_{ie}^{1.5} x_i}{\sum_{i=1}^{4} U_{ie}^{1.5}} = \frac{U_{Ae}^{1.5} x_i + U_{Be}^{1.5} x_i + U_{Ce}^{1.5} x_i + U_{eD}^{1.5} x_i}{U_{Ae}^{1.5} + U_{Be}^{1.5} + U_{eC}^{1.5} + U_{De}^{1.5}}; i = 1, 2$$

ดังนั้นจึงปรับเปลี่ยนตัวแทนกลุ่มแต่ละกลุ่มใหม่ได้ดังนี้

กลุ่ม	ตัวแทนกลุ่ม	
	r_1	r_2
1	1.2495	1.1024
2	3.4905	0.0920

(5) นำค่าของตัวแทนกลุ่มที่ได้ทำการปรับค่าแล้วจากขั้นตอนที่ 4 มาทำซ้ำในขั้นตอนที่ 2 ถึงขั้นตอนที่ 4 กระทั่งข้อมูลไม่มีการเปลี่ยนแปลงกลุ่ม

กลุ่ม	ค่าความเป็นสมาชิก			
	A	B	C	D
1	0.0673	0.9856	0.0453	0.9897
2	0.9327	0.0144	0.9547	0.0103

จากตาราง พบว่าไม่มีการเปลี่ยนแปลงกลุ่มแล้วจึงหยุดดำเนินการ และสรุปได้ว่า สามารถจัดกลุ่มข้อมูลออกเป็น 2 กลุ่ม ดังนี้ กลุ่มที่ 1 คือ (BD) และกลุ่มที่ 2 คือ (AC)

2.7 เทคนิคการจัดกลุ่มข้อมูลแบบอัลกอริทึม EM (Expectation-Maximization Algorithm)

เป็นการจัดกลุ่มข้อมูลที่แต่ละข้อมูลมีการกระจายความน่าจะเป็นในแต่ละกลุ่ม โดยการประมาณค่าเฉลี่ยถ่วงน้ำหนักจากความน่าจะเป็นของข้อมูลในกลุ่มนั้น

อัลกอริทึม EM สำหรับการแจกแจงปกติแบบผสม (EM algorithm for two-component Gaussian mixture) เริ่มต้นด้วยฟังก์ชันการแจกแจงปกติ และให้ X เป็นฟังก์ชันการแจกแจงปกติแบบผสม ซึ่งมีฟังก์ชันดังนี้

$$\begin{aligned} X_1 &\sim N\left(\mu_1, \Sigma_1\right) \\ X_2 &\sim N\left(\mu_2, \Sigma_2\right) \\ X &= \left(1-\Delta\right) X_1 + \Delta X_2 \end{aligned} \tag{2.7.1}$$

โดยที่ $\Delta \in \{0,1\}$ และ $P\left(\Delta = 1\right) = \pi.$ คือการสุ่มที่ชัดเจน เมื่อสุ่ม $\Delta \in \{0,1\}$ ด้วยความน่าจะเป็น π จากนั้นผลที่ได้ออกมาขึ้นอยู่กับ X_1 หรือ X_2 ก็ได้ ให้ $\phi_\theta\left(x\right)$ เป็นความหนาแน่นของการแจกแจงปกติ ที่มีค่าพารามิเตอร์เท่ากับ $\theta = \left(\mu, \Sigma\right)$ จะได้ความหนาแน่นของ X ดังนี้

$$g_X\left(x\right) = \left(1-\pi\right)\phi_{\theta_1}\left(x\right) + \pi\phi_{\theta_2}\left(x\right) \tag{2.7.2}$$

โมเดลที่เหมาะสมโดยใช้ความน่าจะเป็นสูงสุด ซึ่งมีพารามิเตอร์ ดังนี้

$$\theta = \left(\pi, \theta_1, \theta_2\right) = \left(\pi, \mu_1, \Sigma_1, \mu_2, \Sigma_2\right)$$

ฟังก์ชันของ log likelihood คือ

$$\ell\left(\theta; \mathrm{X}\right) = \sum_{r=1}^{p}\sum_{i=1}^{n} \log\left[\left(1-\pi\right)\phi_{\theta_1}\left(x_{ir}\right) + \left(\pi\right)\phi_{\theta_2}\left(x_{ir}\right)\right] \tag{2.7.3}$$

การหาค่าสูงสุดของ $\ell\left(\theta; \mathrm{X}\right)$ ทำได้ยาก เพราะ ผลรวมของเทอมภายในล็อกการิทึม แต่ก็เป็นวิธีการที่เรียบง่าย โดยที่เราต้องพิจารณาตัวแปรแฝงที่ไม่ได้เป็นค่าสังเกต Δ_{ir} มีค่าเป็น 0 หรือ 1 ในสมการที่ (2.7.1) เมื่อ $\Delta_{ir} = 1$ แสดงว่า x_{ij} มาจากกลุ่มที่ 2 แต่ถ้า $\Delta_{ir} = 0$ แสดงว่า x_{ij} มาจากกลุ่มที่ 1 สมมติว่าทราบค่า Δ_{ir} ดังนั้นฟังก์ชันล็อกความน่าจะเป็น คือ

$$\begin{aligned} \ell_0\left(\theta; \mathrm{X}, \Delta\right) = &\sum_{r=1}^{p}\sum_{i=1}^{n}\left[\left(1-\Delta_{ir}\right)\log\phi_{\theta_1}\left(x_{ir}\right) + \Delta_{ir}\log\phi_{\theta_2}\left(x_{ir}\right)\right] \\ &+ \sum_{r=1}^{p}\sum_{i=1}^{n}\left[\left(1-\Delta_{ir}\right)\log\left(1-\pi\right) + \Delta_{ir}\log\pi\right] \end{aligned} \tag{2.7.4}$$

ประมาณค่าความน่าจะเป็นสูงสุดของ μ_1 และ Σ_1 ซึ่งเป็นค่าเฉลี่ยตัวอย่าง และค่าเมทริกซ์ความแปรปรวนสำหรับข้อมูลที่ $\Delta_{ir} = 0$ เช่นเดียวกันกับ μ_2 และ Σ_2 ซึ่งเป็นค่าเฉลี่ยตัวอย่าง และค่าเมท

ริกซ์ความแปรปรวนสำหรับข้อมูลที่ $\Delta_{ir}=1$ การประมาณค่าของ π เป็นค่าสัดส่วนของ $\Delta_{ir}=1$ สามารถหาค่าความน่าจะเป็นสูงสุดสำหรับโมเดลผสมได้ดังนี้

$$\begin{aligned}\frac{\partial \ell}{\partial \theta_2} &= \sum_{r=1}^{p}\sum_{i=1}^{n}\frac{1}{(1-\hat{\pi})\phi_{\hat{\theta}_1}(x_{ir})+\hat{\pi}\phi_{\theta_2}(x_{ir})}\hat{\pi}\frac{\partial \phi_{\theta_2}(x_{ir})}{\partial \theta_2} \\ &= \sum_{r=1}^{p}\sum_{i=1}^{n}\frac{\pi\phi_{\theta_2}(x_{ir})}{(1-\hat{\pi})\phi_{\hat{\theta}_1}(x_{ir})+\hat{\pi}\phi_{\theta_2}(x_{ir})}\frac{1}{\phi_{\theta_2}(x_{ir})}\frac{\partial \phi_{\theta_2}(x_{ir})}{\partial \theta_2} \\ &= \sum_{r=1}^{p}\sum_{i=1}^{n}\frac{\pi\phi_{\theta_2}(x_{ir})}{(1-\hat{\pi})\phi_{\hat{\theta}_1}(x_{ir})+\hat{\pi}\phi_{\theta_2}(x_{ir})}\frac{\partial \log\phi_{\theta_2}(x_{ir})}{\partial \theta_2}\end{aligned}$$

โดยอนุพันธ์ของล็อกความน่าจะเป็น คือ

$$\sum_{r=1}^{p}\sum_{i=1}^{n}\frac{\partial \log\phi_{\theta_2}(x_{ir})}{\partial \theta_2}$$

ดังนั้น ความน่าจะเป็นสูงสุดสำหรับโมเดลผสม คือ การถ่วงน้ำหนักความน่าจะเป็นสูงสุด เมื่อน้ำหนักของ x_{ir} ขึ้นอยู่กับกลุ่ม ดังสมการ

$$\hat{\gamma}_{ir}=\frac{\hat{\pi}\phi_{\hat{\theta}_2}(x_{ir})}{(1-\hat{\pi})\phi_{\hat{\theta}_1}(x_{ir})+\hat{\pi}\phi_{\hat{\theta}_2}(x_{ir})} \qquad (2.7.5)$$

โดยจะในใช้วิธีการที่เรียกว่า อัลกอริทึม EM ซึ่งอยู่ในขั้นตอนการหาค่าคาดหวัง (Expectation step; E-step) เป็นการประมาณค่าพารามิเตอร์ปัจจุบัน เพื่อใช้กำหนดข้อจำกัดให้เป็นไปตามความหนาแน่นของแต่ละโมเดล ส่วนขั้นตอนการหาค่าสูงสุด (Maximization step; M-step) ใช้การถ่วงน้ำหนักความน่าจะเป็นสูงสุด สามารถหาค่าพารามิเตอร์ได้ดังนี้

$$\frac{\partial \ell}{\partial \mu_{1r}}=0\Rightarrow \mu_{1r}=\frac{\sum_{r=1}^{p}\sum_{i=1}^{n}(1-\hat{\gamma}_{ir})x_{ir}}{\sum_{r=1}^{p}\sum_{i=1}^{n}(1-\hat{\gamma}_{ir})} \qquad (2.7.6)$$

$$\frac{\partial \ell}{\partial \mu_{2r}}=0\Rightarrow \mu_{2r}=\frac{\sum_{r=1}^{p}\sum_{i=1}^{n}\hat{\gamma}_{ir}x_{ir}}{\sum_{r=1}^{p}\sum_{i=1}^{N}\hat{\gamma}_{ir}(x_{ir})} \qquad (2.7.7)$$

ในการกำหนดค่า $\hat{\mu}_{1r}$ และ $\hat{\mu}_{2r}$ สามารถเลือกจากค่า x_{ir} ได้ และสัดส่วนผสม $\hat{\pi}$ สามารถเริ่มต้นที่ค่า 0.5 ดังนั้นวิธีอัลกอริทึม EM สามารถสรุปขั้นตอนได้ ดังนี้

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

(1) เริ่มต้นด้วยการกำหนดค่าพารามิเตอร์เริ่มต้น $\left(\hat{\pi}, \mu_1, \Sigma_1, \mu_2, \Sigma_2\right)$

(2) ขั้นตอนการหาค่าคาดหวัง (E): คำนวณ

$$\hat{\gamma}_{ir} = \frac{\hat{\pi}\phi_{\hat{\theta}_2}\left(x_{ir}\right)}{\left(1-\hat{\pi}\right)\phi_{\hat{\theta}_1}\left(x_{ir}\right)+\hat{\pi}\phi_{\hat{\theta}_2}\left(x_{ir}\right)}; \begin{cases} i = 1,2,\ldots \mathrm{n} \\ r = 1,2,\ldots \mathrm{p} \end{cases}$$

(3) ขั้นตอนการหาค่าสูงสุด (M): คำนวณ

$$\mu_{1r} = \frac{\sum_{r=1}^{p}\sum_{i=1}^{n}\left(1-\hat{\gamma}_{ir}\right)x_{ir}}{\sum_{r=1}^{p}\sum_{i=1}^{n}\left(1-\hat{\gamma}_{ir}\right)}, \ \mu_{2r} = \frac{\sum_{r=1}^{p}\sum_{i=1}^{n}\hat{\gamma}_{ir}x_{ir}}{\sum_{r=1}^{p}\sum_{i=1}^{N}\hat{\gamma}_{ir}\left(x_{ir}\right)}$$

มีค่าสัดส่วนผสม เท่ากับ $\hat{\pi} = \frac{1}{nr}\sum_{r=1}^{p}\sum_{i=1}^{n}\hat{\gamma}_{ir}$

(4) ทำขั้นตอนที่ (2) และ (3) ซ้ำ จนกระทั่งลู่เข้า

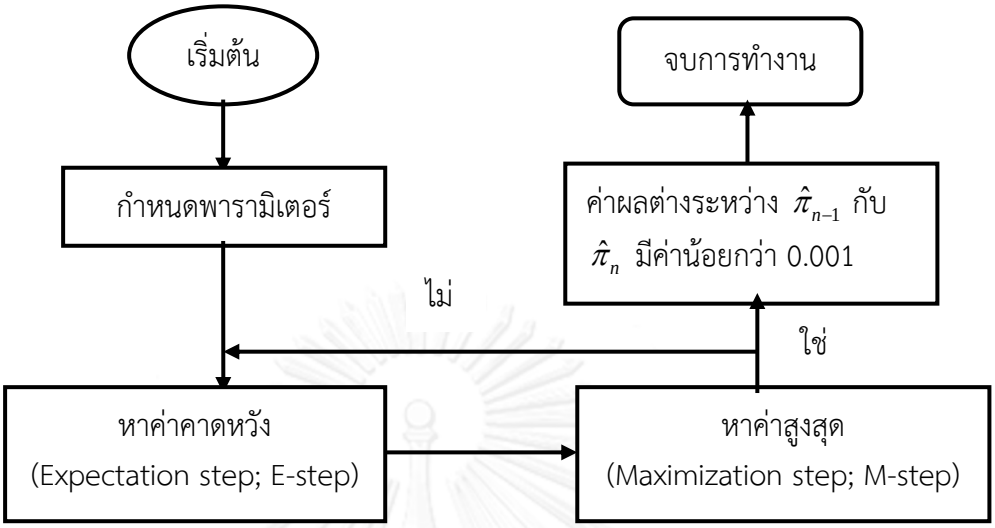
ข้อดีและข้อเสียของวิธีการจัดกลุ่มแบบอัลกอริทึม EM

1. ข้อดี
   - ไม่ได้รับผลกระทบจากค่าผิดปกติ (outlier)
   - ลักษณะของข้อมูลสามารถให้ค่าพารามิเตอร์ที่แตกต่างกัน

2. ข้อเสีย
   - มีความซับซ้อนสูง และใช้เวลาในการคำนวณมาก
   - ต้องทราบรูปแบบของความหนาแน่น

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ขั้นตอนการจัดกลุ่มข้อมูลแบบอัลกอริทึม EM



ภาพที่ 2.7.1 ขั้นตอนการจัดกลุ่มข้อมูลแบบอัลกอริทึม EM

ตัวอย่างที่ 2.7.1 วิธีจัดกลุ่มข้อมูลแบบอัลกอริทึม EM โดยสมมติข้อมูล 20 รายการ ดังนี้ -0.39, 0.12, 0.94, 1.67, 1.76, 2.44, 3.72, 4.28, 4.92, 5.53, 0.06, 0.48, 1.01, 1.68, 1.80, 3.25, 4.12, 4.60, 5.28, 6.22 ทำการแบ่งกลุ่มเป็น 2 กลุ่ม

(1) กำหนดค่าพารามิเตอร์เริ่มต้น

(2) คำนวณหาค่าคาดหวัง จากสูตรเพื่อนำไปใช้หาในขั้นตอนต่อไป

$$\hat{\gamma}_i = \frac{\hat{\pi}\phi_{\hat{\theta}_2}(x_i)}{(1-\hat{\pi})\phi_{\hat{\theta}_1}(x_i)+\hat{\pi}\phi_{\hat{\theta}_2}(x_i)}$$

(3) คำนวณหาค่าสูงสุด คือ คำนวณหาค่าเฉลี่ย ความแปรปรวน และค่าสัดส่วนผสม $(\hat{\pi})$

$$\hat{\mu}_1 = \frac{\sum_{i=1}^{n}(1-\hat{\gamma}_i)x_i}{\sum_{i=1}^{n}(1-\hat{\gamma}_i)}, \hat{\sigma}_1^2 = \frac{\sum_{i=1}^{n}(1-\hat{\gamma}_i)(x_i-\hat{\mu}_1)^2}{\sum_{i=1}^{n}(1-\hat{\gamma}_i)},$$

$$\hat{\mu}_2 = \frac{\sum_{i=1}^{n}\hat{\gamma}_i x_i}{\sum_{i=1}^{n}\hat{\gamma}_i}, \hat{\sigma}_2^2 = \frac{\sum_{i=1}^{n}\hat{\gamma}_i(x_i-\hat{\mu}_1)^2}{\sum_{i=1}^{n}\hat{\gamma}_i}$$

มีค่าสัดส่วนผสม เท่ากับ $\hat{\pi} = \frac{1}{n}\sum_{i=1}^{n}\hat{\gamma}_i(x_i)$

(4) เมื่อได้ค่าจากขั้นตอนที่ 3 แล้ว นำไปใช้หาในขั้นตอนที่ 2 ซึ่งเป็นการหารอบใหม่ ทำซ้ำไปเรื่อย ๆ จนกระทั่งค่าผลต่างระหว่าง $\hat{\pi}_{n-1}$ กับ $\hat{\pi}_n$ มีค่าน้อยกว่า 0.001 จะได้ค่า $\hat{\pi}$ ดังตาราง

รอบที่	$\hat{\pi}$
1	0.485
5	0.493
10	0.523
15	0.544
20	0.546

เมื่อใช้วิธีจัดกลุ่มข้อมูลแบบอัลกอริทึม EM ทำซ้ำ 20 รอบ ทำให้ได้ค่า $\hat{\pi} = 0.546$ และมีค่าประมาณของตัวพารามิเตอร์ดังนี้

$$\hat{\mu}_1 = 4.62 \qquad \hat{\sigma}_1^2 = 0.87$$

$$\hat{\mu}_2 = 1.06 \qquad \hat{\sigma}_2^2 = 0.77$$

2.8 การวัดประสิทธิภาพการจัดกลุ่มข้อมูล

2.8.1 วิธี Calinski and Harabasz index (Pseudo F)

คือ ค่าอัตราส่วนของความแปรปรวนระหว่างกลุ่มกับความแปรปรวนภายในกลุ่ม เมื่อค่า Pseudo F มีค่ามาก แสดงถึง การจัดกลุ่มที่ดี เกณฑ์นี้จะคล้ายกับ F-test ในการวิเคราะห์ความแปรปรวน (Calinski and Harabasz 1974)

$$PseudoF = \frac{(SSb)/(k-1)}{(SSw)/(n-k)}$$

เมื่อ $SSt = \sum_{z=1}^{k}\sum_{r=1}^{p}(x_{zr} - \bar{x}_r)^2$, $SSW = \sum_{z=1}^{k}\sum_{r=1}^{p}(x_{zr} - \bar{x}_{zr})^2$ และ $SSb = SSt - SSw$

โดยที่ SSt คือ ผลรวมของผลต่างกำลังสองของข้อมูลทั้งหมด

SSb คือ ผลรวมของผลต่างกำลังสองของข้อมูลระหว่างกลุ่ม

SSw คือ ผลรวมของผลต่างกำลังสองของข้อมูลภายในกลุ่ม

x_{zr} คือ ข้อมูลกลุ่มที่ z ตัวแปรศึกษาที่ r

$\bar{x}_r$ คือ ค่าเฉลี่ยข้อมูลในตัวแปรศึกษาที่ r

$\bar{x}_{zr}$ คือ ค่าเฉลี่ยของข้อมูลกลุ่มที่ z ตัวแปรศึกษาที่ r

k คือ จำนวนกลุ่มที่ทำการจัดกลุ่ม

n คือ จำนวนข้อมูลทั้งหมด

2.8.2 วิธี Silhouette width

คือ ค่าที่วัดจากค่าความแตกต่างระหว่างความหนาแน่นภายในกลุ่ม กับส่วนที่เหลือ ซึ่งค่า Silhouette width หรือ $S(i)$ มีค่าอยู่ในช่วง [-1,1] ถ้า $S(i)$ มีค่าใกล้ 1 แสดงว่า ข้อมูลที่ถูกนำมาทำการจัดกลุ่มไว้มีความเหมาะสม แต่ถ้า $S(i)$ มีค่าใกล้ -1 แสดงว่า ข้อมูลที่นำมาทำการจัดกลุ่มไม่เหมาะสมกับกลุ่มที่ถูกจัดไว้ วิธีนี้ถูกเสนอโดย Peter J. Rousseeuw ในปี 1986

$$S_i = \frac{F_i - G_i}{\max(F_i, G_i)}$$

โดยที่ G_i คือ ค่าระยะทางเฉลี่ยของข้อมูลตัวที่ i กับข้อมูลในกลุ่มอื่น

F_i คือ ค่าระยะทางเฉลี่ยของข้อมูลตัวที่ i กับข้อมูลทั้งหมด

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทที่ 3

วิธีการดำเนินงานวิจัย

ในการศึกษาครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพการจัดกลุ่ม สำหรับข้อมูลที่มีการแจกแจงปกติแบบผสม ซึ่งได้เสนอไว้ 4 วิธี คือ วิธีการจัดกลุ่มแบบลำดับชั้น, วิธีการจัดกลุ่มแบบเคมีน, วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน และวิธีการจัดกลุ่มแบบอัลกอริทึม EM โดยทำการศึกษาข้อมูลที่มีการแจกแจงปกติแบบผสม มีจำนวนการซ้อนทับกัน 2, 3 และ 4 กลุ่ม ขนาดตัวอย่างที่ศึกษากลุ่มละ 50, 100 และ 300 ตัวอย่าง ผู้วิจัยทำการจำลองข้อมูลเป็นกลุ่มสำหรับแต่ละขนาดตัวอย่าง จำนวนกลุ่มที่ทำการจัดกลุ่มข้อมูลมีขนาดเป็น 2, 3 และ 4 กลุ่ม ใช้วิธี Pseudo F และวิธี Silhouette width เป็นดัชนีการตัดสินใจในแต่ละวิธีการจัดกลุ่ม ทำการเปรียบเทียบประสิทธิภาพของการจัดกลุ่มข้อมูล โดยพิจารณาจากกรณีที่จำนวนกลุ่มของการจัดกลุ่มข้อมูล เท่ากับจำนวนกลุ่มของการซ้อนทับ สำหรับการศึกษาครั้งนี้ใช้โปรแกรม R เวอร์ชั่น 3.0.2 ซึ่งมีแผนการจำลองข้อมูลและขั้นตอนในการวิจัยดังนี้

3.1 ขอบเขตของการดำเนินการวิจัย

ในการวิจัยครั้งนี้จะทำการศึกษาภายใต้ขอบเขตดังนี้

1. จำลองข้อมูลที่มีการแจกแจงปกติแบบผสม

2. ข้อมูลที่ถูกจำลองมีอัตราการซ้อนทับเฉลี่ย $(\bar{\omega})$ ที่ 1%, 10%, 20%, 30% และ 40% (Clusters with 1%, 10%, 20%, 30% and 40% Overlapping of average$(\bar{\omega})$)

3. กำหนดจำนวนกลุ่มของการซ้อนทับกัน คือ 2, 3 และ 4 กลุ่ม ส่วนจำนวนตัวแปรที่ใช้ศึกษา คือ 2 และ 3 ตัวแปร โดยจำลองข้อมูลแต่ละกลุ่มขนาด 50, 100 และ 300

4. กำหนดจำนวนกลุ่มของการจัดกลุ่มข้อมูลเท่ากับ 2, 3 และ 4 กลุ่ม

5. ใช้วิธี Pseudo F และวิธี Silhouette width เป็นดัชนีการตัดสินใจ

3.2 ขั้นตอนในการดำเนินการวิจัย

1. จำลองข้อมูลที่ลักษณะการแจกแจงปกติแบบผสม โดยกำหนดค่าการซ้อนทับเฉลี่ยเท่ากับ 1%, 10%, 20%, 30% และ 40%

2. จำลองข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน คือ 2, 3 และ 4 กลุ่ม ตัวแปรที่ใช้ในการศึกษา คือ 2 และ 3 ตัวแปร และขนาดของข้อมูลแต่ละกลุ่มเท่ากับ 50, 100 และ 300 โดยใช้ฟังก์ชัน MixSim ในโปรแกรม R แบ่งการจำลองออกเป็น 2 กรณี คือ

2.1 กรณีที่ฐานข้อมูลอยู่ในรูปแบบวงรี (Non Spherical)

2.2 กรณีที่ฐานข้อมูลอยู่ในรูปแบบวงกลม (Spherical or Isotopic)

3. จัดกลุ่มข้อมูลด้วยวิธีการจัดกลุ่มแบบลำดับชั้น, วิธีการจัดกลุ่มแบบเคมีน, วิธีการจัดกลุ่มแบบฟัซซี่ซี-มีน และวิธีการจัดกลุ่มแบบอัลกอริทึม EM โดยมีจำนวนกลุ่มข้อมูลเท่ากับ 2, 3 และ 4 กลุ่ม

4. ใช้วิธี Pseudo F และวิธี Silhouette width เป็นดัชนีการตัดสินใจในแต่ละวิธีการจัดกลุ่มว่าข้อมูลควรมีจำนวนกลุ่มของการจัดกลุ่มข้อมูลเท่าไหร่

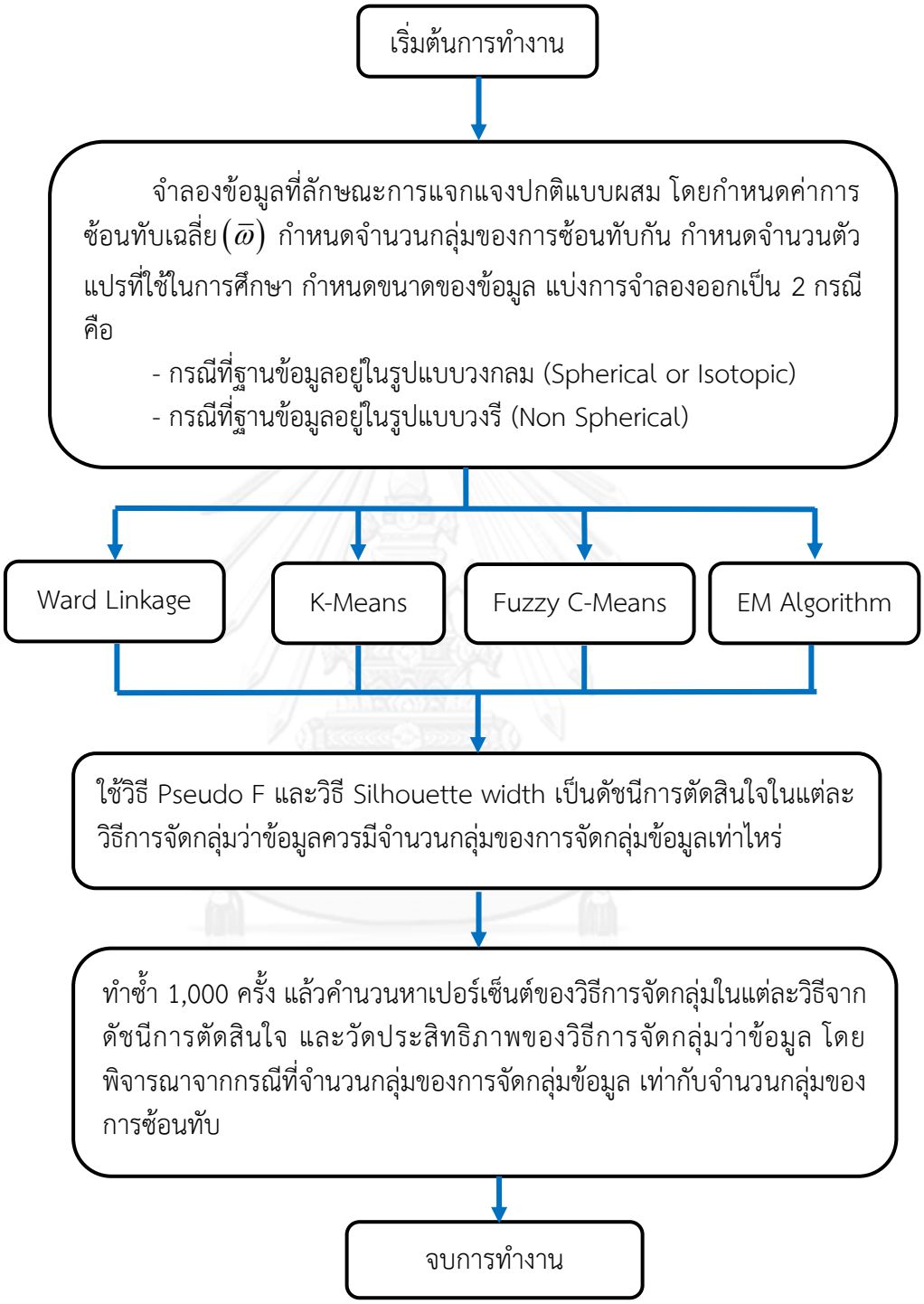
5. ทำซ้ำ 1,000 ครั้ง และคำนวนหาเปอร์เซ็นต์ของวิธีการจัดกลุ่มในแต่ละวิธีจากดัชนีการตัดสินใจ และวัดประสิทธิภาพของวิธีการจัดกลุ่มว่าข้อมูล โดยพิจารณาจากกรณีที่จำนวนกลุ่มของการจัดกลุ่มข้อมูล เท่ากับจำนวนกลุ่มของการซ้อนทับ

6. วัดประสิทธิภาพของวิธีการจัดกลุ่มว่าข้อมูล โดยพิจารณาจากกรณีที่จำนวนกลุ่มของการจัดกลุ่มข้อมูล เท่ากับจำนวนกลุ่มของการซ้อนทับ

7. นำข้อมูลจริงมาศึกษาตามขั้นตอนที่ 3 และขั้นตอนที่ 4 จากนั้นจึงนำมาเปรียบเทียบกับข้อมูลจำลอง และสรุปผล

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

3.3 ขั้นตอนการทำงานของโปรแกรม



บทที่ 4

ผลการวิเคราะห์ข้อมูล

จากการศึกษาเปรียบเทียบประสิทธิภาพการจัดกลุ่มข้อมูล โดยใช้วิธีการจัดกลุ่มข้อมูล 4 วิธี ได้แก่ วิธีการจัดกลุ่มแบบลำดับชั้น, วิธีการจัดกลุ่มแบบเคมีน, วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน และวิธีการจัดกลุ่มแบบอัลกอริทึม EM ทำการจำลองข้อมูลที่มีการแจกแจงปกติแบบผสม โดยแบ่งออกเป็น 2 กรณี ดังนี้ 1.กรณีที่ฐานข้อมูลอยู่ในรูปแบบวงรี (Non Spherical) 2.กรณีที่ฐานข้อมูลอยู่ในรูปแบบวงกลม (Spherical or Isotopic) ทำการจำลองข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 2, 3 และ 4 กลุ่ม ตัวแปรที่ใช้ในการศึกษามี 2 และ 3 ตัวแปร ขนาดข้อมูลแต่ละกลุ่มเท่ากับ 50, 100 และ 300 ตัวอย่าง กำหนดจำนวนกลุ่มของการจัดกลุ่มข้อมูลเท่ากับ 2, 3 และ 4 กลุ่ม ใช้วิธี Calinski and Harabasz index (Pseudo F) และวิธี Silhouette width เป็นเกณฑ์ในการวัดประสิทธิภาพของวิธีการจัดกลุ่มข้อมูล

ในการนำเสนอผลงานวิจัยจะแสดงในรูปแบบตาราง มีสัญลักษณ์ที่ใช้ ดังนี้

$\bar{\omega}$	แทน	ค่าการซ้อนทับเฉลี่ย
k	แทน	จำนวนกลุ่มของการจัดกลุ่มข้อมูล
HC	แทน	วิธีการจัดกลุ่มแบบลำดับชั้น ด้วยวิธี Ward
KC	แทน	วิธีการจัดกลุ่มแบบเคมีน
FC	แทน	วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน
EM	แทน	วิธีการจัดกลุ่มแบบอัลกอริทึม EM

<u>หมายเหตุ</u>

ตัวอักษรสีแดง แทน เปอร์เซ็นต์ของจำนวนกลุ่มที่ค่าดัชนีการตัดสินใจดีที่สุดในแต่ละรอบ

ช่องแถบสี แทน เปอร์เซ็นต์ของดัชนีการตัดสินใจที่ดีที่สุด พิจารณาจากกรณีที่จำนวนกลุ่มของการจัดกลุ่มข้อมูล เท่ากับจำนวนกลุ่มของการซ้อนทับ

4.1 กรณีที่ฐานข้อมูลอยู่ในรูปแบบวงรี (Non Spherical)

4.1.1 วิธี Calinski and Harabasz index (Pseudo F)

ตารางที่ 4.1.1 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 50 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
2	2	1%	2	72.0	74.3	72.3	82.3
			3	18.4	17.7	18.4	9.3
			4	9.6	8.0	9.3	8.4
		10%	2	46.5	54.8	53.3	51.5
			3	28.3	26.0	25.1	20.4
			4	25.2	19.2	21.6	28.1
		20%	2	39.0	43.6	43.2	30.6
			3	31.3	32.2	30.3	30.1
			4	29.7	24.2	26.5	39.3
		30%	2	35.5	41.8	41.7	27.6
			3	33.5	34.1	33.6	30.0
			4	31.0	24.1	24.7	42.4
		40%	2	33.9	39.7	39.7	22.7
			3	33.2	32.3	33.0	33.8
			4	32.9	28.0	27.3	43.5

จากตารางที่ 4.1.1 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 2 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 10%, 20% และ 30% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 2 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบเคมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 2 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบเคมีน และ วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.1.2 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 100 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
2	2	1%	2	77.4	78.7	77.5	92.7
			3	16.5	16.9	16.6	5.0
			4	6.1	4.4	5.9	2.3
		10%	2	49.7	58.3	58.9	64.1
			3	28.6	24.3	23.1	15.8
			4	21.7	17.4	18.0	20.1
		20%	2	45.4	51.2	51.0	42.1
			3	30.5	31.2	30.9	22.8
			4	24.1	17.6	18.1	35.1
		30%	2	41.1	42.7	42.9	29.9
			3	32.0	37.3	39.2	27.7
			4	26.9	20.0	17.9	42.4
		40%	2	41.2	43.6	44.3	18.6
			3	31.7	33.2	33.2	31.8
			4	27.1	23.2	22.5	49.6

จากตารางที่ 4.1.2 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1% และ 10% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 2 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 20% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 2 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบเคมีนเป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 30% และ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 2 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.1.3 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ดีที่สุดในแต่ละรอบ สำหรับข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 300 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
2	2	1%	2	81.8	81.5	81.0	93.3
			3	14.6	14.5	14.6	5.2
			4	3.6	4.0	4.4	1.5
		10%	2	53.3	59.6	60.0	75.7
			3	29.0	24.5	25.2	12.5
			4	17.7	15.9	14.8	11.8
		20%	2	46.7	51.9	52.6	58.8
			3	31.7	35.0	35.3	20.4
			4	21.6	13.1	12.1	20.8
		30%	2	39.1	45.5	45.5	41.6
			3	37.8	41.0	42.1	27.3
			4	23.1	13.5	12.4	31.1
		40%	2	43.9	49.3	49.9	26.4
			3	33.8	34.1	35.6	27.7
			4	22.3	16.6	14.5	45.9

จากตารางที่ 4.1.3 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1%, 10% และ 20% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 2 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 30% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 2 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบเคมีนและ วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 2 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.1.4 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 50 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
2	3	1%	2	96.2	96.3	96.3	96.5
			3	3.3	3.3	3.6	3.1
			4	0.5	0.4	0.1	0.4
		10%	2	65.1	71.0	71.0	58.9
			3	23.5	20.9	21.1	24.1
			4	11.4	8.1	7.9	17.0
		20%	2	55.3	59.3	59.1	37.0
			3	30.2	31.4	31.5	34.9
			4	14.5	9.3	9.4	28.1
		30%	2	48.5	56.5	56.4	31.7
			3	31.1	30.8	32.2	35.0
			4	20.4	12.7	11.4	33.3
		40%	2	51.3	61.8	62.8	32.7
			3	29.1	27.7	28.4	34.3
			4	19.6	10.5	8.8	33.0

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จากตารางที่ 4.1.4 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 2 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 10% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 2 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบเคมีน และ วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 20% และ 30% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 2 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบเคมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 2 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 4.1.5 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 100 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
2	3	1%	2	98.0	97.5	97.3	99.0
			3	1.8	2.5	2.7	0.7
			4	0.2	0	0	0.3
		10%	2	68.7	75.0	75.1	71.7
			3	22.5	19.9	19.9	13.4
			4	8.8	5.1	5.0	14.9
		20%	2	54.6	58.1	58.6	41.2
			3	31.2	34.6	33.6	28.1
			4	14.2	7.3	7.8	30.7
		30%	2	53.7	60.1	60.2	25.6
			3	32.5	32.6	33.9	37.5
			4	13.8	7.3	5.9	36.9
		40%	2	56.4	64.5	65.8	25.3
			3	28.4	28.6	29.1	32.9
			4	15.2	6.9	5.1	41.8

จากตารางที่ 4.1.5 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 2 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 10%, 20%, 30% และ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 2 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.1.6 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 300 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
2	3	1%	2	98.7	97.9	97.9	99.6
			3	1.3	2.1	2.1	0.4
			4	0	0	0	0
		10%	2	75.8	78.6	79.0	91.2
			3	18.2	17.6	16.9	5.0
			4	6.0	3.8	4.1	3.8
		20%	2	57.8	64.4	64.4	68.1
			3	30.2	29.5	29.4	15.5
			4	12.0	6.1	6.2	16.4
		30%	2	58.3	63.2	62.2	41.0
			3	29.1	30.2	31.2	22.8
			4	12.6	6.6	6.6	36.2
		40%	2	59.5	67.9	68.8	19.5
			3	29.0	27.1	27.1	28.6
			4	11.5	5.0	4.1	51.9

จากตารางที่ 4.1.6 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1%, 10% และ 20% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 2 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 30% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 2 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบเคมีนเป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 2 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.1.7 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 50 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
3	2	1%	2	15.0	16.4	14.6	18.1
			3	59.4	60.3	63.2	69.8
			4	25.6	23.3	22.2	12.1
		10%	2	46.5	45.8	45.2	53.1
			3	28.9	33.0	32.8	29.4
			4	24.6	21.2	22.0	17.5
		20%	2	40.9	42.8	43.3	40.3
			3	27.6	29.7	28.9	28.0
			4	31.5	27.5	27.8	31.7
		30%	2	34.1	35.5	36.4	27.8
			3	29.7	32.9	32.6	26.8
			4	36.2	31.6	31.0	45.4
		40%	2	33.5	32.9	34.9	18.2
			3	29.2	30.7	33.1	29.6
			4	37.3	36.4	32.0	52.2

จากตารางที่ 4.1.7 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 3 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 10%, 20% และ 30% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 3 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบเคมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 3 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบฟัซซี่ ซีมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.1.8 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ดีที่สุดในแต่ละรอบ สำหรับข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 100 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
3	2	1%	2	12.9	13.4	12.4	15.6
			3	64.6	65.8	67.5	77.4
			4	22.5	20.8	20.1	7.0
		10%	2	48.9	48.7	48.6	55.8
			3	28.4	30.6	31.8	32.3
			4	22.7	20.7	19.6	11.9
		20%	2	43.6	46.5	46.3	49.9
			3	29.3	28.3	29.1	26.2
			4	27.1	25.2	24.6	23.9
		30%	2	36.1	40.9	41.8	37.5
			3	32.0	31.2	31.4	24.2
			4	31.9	27.9	26.8	38.3
		40%	2	38.4	40.6	41.3	21.9
			3	30.1	30.1	31.2	26.4
			4	31.5	29.3	27.5	51.7

จากตารางที่ 4.1.8 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1% และ 10% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 3 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 20% และ 30% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 3 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบลำดับชั้น เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 3 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบฟัซซี่ ซีมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.1.9 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 300 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
3	2	1%	2	12.6	13.2	12.9	15.0
			3	68.0	64.6	68.1	79.5
			4	19.4	22.2	19.0	5.5
		10%	2	49.0	48.7	48.8	56.0
			3	28.7	33.1	33.2	36.7
			4	22.3	18.2	18.0	7.3
		20%	2	46.2	46.1	46.3	58.3
			3	27.8	30.3	29.9	25.7
			4	26.0	23.6	23.8	16.0
		30%	2	41.7	41.6	41.4	45.6
			3	30.6	30.6	33.1	26.0
			4	27.7	27.8	25.5	28.4
		40%	2	40.4	40.5	42.3	29.7
			3	30.9	30.8	32.2	30.2
			4	28.7	28.7	25.4	40.1

จากตารางที่ 4.1.9 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1% และ 10% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 3 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 20% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 3 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบเคมีนเป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 30% และ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 3 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.1.10 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 50 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
3	3	1%	2	23.4	25.8	22.8	23.2
			3	72.3	66.2	73.8	73.2
			4	4.3	8.0	3.4	3.6
		10%	2	56.9	59.3	58.3	55.9
			3	30.1	33.6	34.9	26.6
			4	13.0	7.1	6.8	17.5
		20%	2	54.6	59.9	60.4	45.1
			3	25.6	27.6	28.3	25.0
			4	19.8	12.5	11.3	29.9
		30%	2	46.5	51.8	53.2	30.2
			3	30.1	32.3	31.9	33.4
			4	23.4	15.9	14.9	36.4
		40%	2	51.2	61.3	62.3	29.3
			3	25.6	25.9	28.2	34.5
			4	23.2	12.8	9.5	36.2

จากตารางที่ 4.1.10 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1%, 10% และ 20% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลป็น 3 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 30% และ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลป็น 3 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.1.11 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 100 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
3	3	1%	2	21.2	23.3	20.6	21.5
			3	75.7	71.7	77.3	77.8
			4	3.1	5.0	2.1	0.7
		10%	2	62.1	62.3	61.9	62.6
			3	27.6	31.0	31.8	28.6
			4	10.3	6.7	6.3	8.8
		20%	2	59.9	63.6	64.4	54.8
			3	25.1	26.7	26.8	20.4
			4	15.0	9.7	8.8	24.8
		30%	2	50.8	56.3	56.5	33.4
			3	30.6	30.4	31.9	27.1
			4	18.6	13.3	11.6	39.5
		40%	2	53.0	60.7	62.0	24.3
			3	28.3	28.6	30.0	32.1
			4	18.7	10.7	8.0	43.6

จากตารางที่ 4.1.11 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1% และ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 3 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 10%, 20% และ 30% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 3 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.1.12 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 300 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
3	3	1%	2	23.0	24.6	20.4	22.3
			3	75.6	70.5	77.8	77.4
			4	1.4	4.9	1.8	0.3
		10%	2	63.1	63.4	63.3	63.5
			3	26.0	29.4	29.4	35.0
			4	10.9	7.2	7.3	1.5
		20%	2	63.2	63.5	63.6	66.1
			3	22.8	26.0	26.8	26.5
			4	14.0	10.5	9.7	7.4
		30%	2	54.8	57.9	59.0	54.6
			3	27.1	31.7	31.4	23.8
			4	18.1	10.4	10.4	9.6
		40%	2	58.8	66.0	66.3	30.7
			3	28.3	28.3	29.0	23.7
			4	12.9	5.7	4.7	45.6

จากตารางที่ 4.1.12 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1%, 20% และ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 3 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 10% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 3 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 30% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 3 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบเคมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.1.13 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 50 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
4	2	1%	2	3.1	8.8	4.3	3.1
			3	13.6	31.6	23.5	17.6
			4	83.3	59.6	72.2	79.3
		10%	2	27.9	28.2	28.0	32.5
			3	37.9	40.4	39.9	41.7
			4	34.2	31.4	32.1	25.8
		20%	2	36.0	37.2	37.2	41.7
			3	30.3	33.8	33.6	25.8
			4	33.7	29.0	29.2	32.5
		30%	2	32.5	33.1	33.3	27.5
			3	30.4	32.8	33.6	26.2
			4	37.1	34.1	33.1	46.3
		40%	2	34.6	34.9	36.0	19.1
			3	27.7	29.3	31.4	27.6
			4	37.7	35.8	32.6	53.3

จากตารางที่ 4.1.13 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1%, 10% และ 20% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 4 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบลำดับชั้น เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 30% และ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 4 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.1.14 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 100 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
4	2	1%	2	1.6	7.7	3.6	2.1
			3	15.2	32.3	25.9	17.5
			4	83.2	60.0	70.5	80.4
		10%	2	28.2	26.0	24.9	37.3
			3	37.2	40.8	40.8	38.6
			4	34.6	33.2	34.3	24.1
		20%	2	38.2	38.0	38.4	43.8
			3	29.5	32.9	32.0	30.7
			4	32.2	29.1	29.6	25.5
		30%	2	36.0	36.2	37.8	34.7
			3	31.2	31.2	32.7	26.8
			4	32.8	32.6	29.5	38.5
		40%	2	37.1	37.4	38.6	19.3
			3	28.5	29.2	32.1	27.8
			4	34.4	33.4	29.3	52.9

จากตารางที่ 4.1.14 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1%, 10% และ 20% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 4 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบลำดับชั้น เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 30% และ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 4 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.1.15 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 300 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
4	2	1%	2	2.4	5.9	3.3	2.5
			3	16.2	34.0	25.3	16.9
			4	81.4	60.1	71.4	80.6
		10%	2	31.0	27.9	27.6	33.4
			3	38.5	40.8	41.0	42.0
			4	30.5	31.3	31.4	24.6
		20%	2	43.8	43.6	43.6	48.1
			3	27.6	31.2	32.2	31.0
			4	28.6	25.2	24.2	20.9
		30%	2	37.7	36.9	38.1	41.5
			3	30.7	34.4	34.6	29.7
			4	31.6	28.7	27.3	28.8
		40%	2	39.4	36.6	39.3	27.3
			3	27.4	32.0	33.3	28.7
			4	33.2	31.4	27.4	44.0

จากตารางที่ 4.1.15 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1%, 20% และ 30% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 4 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบลำดับชั้น เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 10% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 4 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 4 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.1.16 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 50 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
4	3	1%	2	5.6	11.0	5.3	6.1
			3	19.9	25.7	20.6	19.3
			4	74.5	63.3	74.1	74.6
		10%	2	43.9	45.0	44.5	43.5
			3	35.9	38.3	39.4	35.1
			4	20.2	16.7	16.1	21.4
		20%	2	51.6	54.1	53.7	42.1
			3	26.3	28.5	29.6	27.0
			4	22.1	17.4	16.7	30.9
		30%	2	45.9	53.0	55.2	30.3
			3	31.4	32.5	31.5	33.2
			4	22.7	14.5	13.3	36.5
		40%	2	48.8	63.0	64.9	27.4
			3	27.9	25.3	25.8	34.1
			4	23.3	11.7	9.3	38.5

จากตารางที่ 4.1.16 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1%, 10%, 20%, 30% และ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 4 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.1.17 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 100 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
4	3	1%	2	6.0	9.9	5.2	5.6
			3	18.0	27.4	19.7	17.6
			4	76.0	62.7	75.1	76.8
		10%	2	43.7	43.4	42.6	46.8
			3	37.7	40.0	41.4	38.6
			4	18.6	16.6	16.0	14.6
		20%	2	53.9	55.8	56.6	46.6
			3	29.5	30.6	30.9	28.9
			4	16.6	13.6	12.5	24.5
		30%	2	49.1	55.0	55.8	33.0
			3	30.5	32.3	32.7	26.5
			4	20.4	12.7	11.5	40.5
		40%	2	51.2	61.8	63.7	18.1
			3	30.3	28.5	28.9	32.1
			4	18.5	9.7	7.4	49.8

จากตารางที่ 4.1.17 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1%, 20%, 30% และ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 4 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 10% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 4 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบลำดับชั้น เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.1.18 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 300 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
4	3	1%	2	5.0	9.0	5.2	4.8
			3	20.5	29.3	21.5	19.7
			4	74.5	61.7	73.3	75.5
		10%	2	48.0	47.3	46.7	47.7
			3	36.3	39.9	40.6	37.8
			4	15.7	12.8	12.7	14.5
		20%	2	55.2	56.6	57.4	56.8
			3	29.1	31.0	31.4	31.1
			4	15.7	12.4	11.2	12.1
		30%	2	55.7	60.3	60.5	45.5
			3	28.5	28.1	28.7	30.8
			4	15.8	11.6	10.8	23.7
		40%	2	58.5	65.1	65.8	27.9
			3	28.4	27.3	28.4	28.1
			4	13.1	7.6	5.8	44.0

จากตารางที่ 4.1.18 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1%, 30% และ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 4 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 10% และ 20% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 4 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบลำดับชั้น เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

4.1.2 วิธี Silhouette width

ตารางที่ 4.1.19 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 50 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
2	2	1%	2	90.4	91.0	92.0	93.9
			3	7.7	7.8	6.8	4.2
			4	1.9	1.2	1.2	1.9
		10%	2	76.9	79.6	80.4	73.8
			3	16.3	15.9	15.5	14.8
			4	6.8	4.5	4.1	11.4
		20%	2	70.5	70.4	70.0	55.5
			3	19.6	21.5	21.8	24.3
			4	9.9	8.1	8.2	20.2
		30%	2	69.0	66.2	66.5	49.7
			3	21.7	23.9	24.5	29.3
			4	9.3	9.9	9.0	21.0
		40%	2	66.3	65.6	62.8	54.4
			3	22.7	22.8	26.4	25.8
			4	11.0	11.6	10.8	19.8

จากตารางที่ 4.1.19 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 2 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 10% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 2 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 20%, 30% และ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 2 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบลำดับชั้น เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.1.20 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 100 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
2	2	1%	2	94.1	94.0	94.7	97.5
			3	5.2	5.5	4.8	2.1
			4	0.7	0.5	0.5	0.4
		10%	2	76.8	78.9	79.8	82.4
			3	18.1	17.2	17.0	11.9
			4	5.1	3.9	3.2	5.7
		20%	2	74.0	72.0	72.7	62.2
			3	18.8	22.9	22.9	23.1
			4	7.2	5.1	4.4	14.7
		30%	2	73.2	70.7	68.5	51.0
			3	19.0	23.9	27.4	26.3
			4	7.8	5.4	4.1	22.7
		40%	2	73.2	68.4	69.3	46.1
			3	19.2	21.9	23.7	27.2
			4	7.6	9.7	7.0	26.7

จากตารางที่ 4.1.20 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1% และ 10% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 2 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 20%, 30% และ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 2 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบลำดับชั้น เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.1.21 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 300 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
2	2	1%	2	93.9	93.2	93.2	97.6
			3	5.7	6.7	6.7	1.9
			4	0.4	0.1	0.1	0.5
		10%	2	79.6	79.9	81.7	89.4
			3	17.1	18.2	16.9	9.0
			4	3.3	1.9	1.4	1.6
		20%	2	78.5	75.5	76.3	77.0
			3	17.8	22.0	22.5	17.9
			4	3.7	2.5	1.2	5.1
		30%	2	74.2	70.5	70.0	61.2
			3	20.6	26.7	28.2	24.6
			4	5.2	2.8	1.8	14.2
		40%	2	73.8	69.4	70.4	50.4
			3	19.7	25.3	25.5	27.6
			4	6.5	5.3	4.1	22.0

จากตารางที่ 4.1.21 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1% และ 10% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 2 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 20%, 30% และ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 2 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบลำดับชั้น เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.1.22 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 50 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
2	3	1%	2	96.9	97.3	97.5	97.5
			3	2.7	2.3	2.4	2.2
			4	0.4	0.4	0.1	0.3
		10%	2	75.6	79.8	82.2	69.5
			3	18.0	14.7	14.4	20.1
			4	6.4	5.5	3.4	10.4
		20%	2	69.4	68.8	69.8	55.8
			3	20.9	22.6	22.3	25.5
			4	9.7	8.6	7.9	18.7
		30%	2	64.0	63.7	65.1	54.6
			3	23.1	22.9	24.5	27.1
			4	12.9	13.4	10.4	18.3
		40%	2	65.4	61.9	65.6	59.6
			3	19.9	23.3	24.6	24.2
			4	14.7	14.8	9.8	16.2

จากตารางที่ 4.1.22 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 2 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบอัลกอริทึม EM และ วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 10%, 30% และ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 2 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 20% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 2 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบลำดับชั้น เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.1.23 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 100 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
2	3	1%	2	98.3	98.5	98.5	99.2
			3	1.5	1.5	1.5	0.8
			4	0.2	0	0	0
		10%	2	80.9	81.1	84.2	79.9
			3	15.6	15.6	13.4	13.5
			4	3.5	3.3	2.4	6.6
		20%	2	71.2	68.6	70.9	57.1
			3	22.1	24.9	23.2	26.5
			4	6.7	6.5	5.9	16.4
		30%	2	72.2	67.8	68.5	45.0
			3	19.7	24.2	24.9	34.3
			4	8.1	8.0	6.6	20.7
		40%	2	71.3	65.0	69.3	44.6
			3	20.8	25.0	24.5	31.4
			4	7.9	10.0	6.2	24.0

จากตารางที่ 4.1.23 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 2 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 10% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 2 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 20%, 30% และ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 2 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบลำดับชั้น เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.1.24 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 300 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
2	3	1%	2	99.0	98.3	98.4	99.8
			3	0.9	1.7	1.6	0.2
			4	0.1	0	0	0
		10%	2	83.1	84.7	87.1	92.4
			3	15.5	13.1	11.5	7.0
			4	1.4	2.2	1.4	0.6
		20%	2	79.1	73.5	74.5	78.5
			3	16.7	21.5	21.4	16.0
			4	4.2	5.0	4.1	5.5
		30%	2	74.6	68.1	68.1	55.5
			3	19.2	24.8	25.4	24.5
			4	6.2	7.1	6.5	20.0
		40%	2	76.0	68.6	71.7	35.3
			3	18.2	22.1	22.1	28.1
			4	5.8	9.3	6.2	36.6

จากตารางที่ 4.1.24 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1% และ 10% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 2 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 20%, 30% และ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 2 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบลำดับชั้น เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.1.25 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 50 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
3	2	1%	2	62.7	63.0	60.6	61.7
			3	35.3	34.9	38.1	36.8
			4	2.0	2.1	1.3	1.5
		10%	2	74.2	71.2	71.2	71.0
			3	18.2	22.9	23.5	18.6
			4	7.6	5.9	5.3	10.4
		20%	2	66.4	63.9	64.9	58.9
			3	24.1	26.1	25.4	23.8
			4	9.5	10.0	9.7	17.3
		30%	2	59.9	57.8	56.7	49.6
			3	26.4	27.5	30.7	25.0
			4	13.7	14.7	12.6	25.4
		40%	2	59.8	55.9	54.5	48.6
			3	27.4	29.4	32.5	28.2
			4	12.8	14.7	13.0	23.2

จากตารางที่ 4.1.25 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1%, 10%, 30% และ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 3 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 20% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 3 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบเคมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.1.26 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 100 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
3	2	1%	2	60.2	58.7	56.3	58.5
			3	37.7	38.8	42.1	40.3
			4	2.1	2.5	1.6	1.2
		10%	2	74.2	71.6	71.8	73.7
			3	19.1	23.4	23.8	20.7
			4	6.7	5.0	4.4	5.6
		20%	2	69.6	67.8	68.5	63.9
			3	21.7	24.6	25.4	22.4
			4	8.7	7.6	6.1	13.7
		30%	2	65.9	62.8	63.9	56.9
			3	23.4	27.5	28.5	21.3
			4	10.7	9.7	7.6	21.8
		40%	2	66.5	60.2	59.3	47.5
			3	22.1	26.1	29.2	28.7
			4	11.4	13.7	11.5	23.8

จากตารางที่ 4.1.26 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1%, 10%, 20%, 30% และ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 3 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบฟัซซีซีมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.1.27 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 300 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
3	2	1%	2	62.5	61.0	58.2	61.1
			3	35.4	37.0	40.5	37.7
			4	2.1	2.0	1.3	1.2
		10%	2	75.3	72.4	73.2	73.2
			3	19.7	22.8	22.7	22.5
			4	5.0	4.8	4.1	4.3
		20%	2	70.5	67.8	67.8	68.6
			3	21.9	24.9	25.9	20.6
			4	7.6	7.3	6.3	10.8
		30%	2	71.3	63.5	62.6	61.7
			3	22.5	26.7	29.5	22.0
			4	6.2	9.8	7.9	16.3
		40%	2	68.5	59.0	59.7	54.5
			3	22.7	30.3	31.9	26.6
			4	8.8	10.7	8.4	18.9

จากตารางที่ 4.1.27 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1%, 20%, 30% และ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 3 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 10% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 3 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบเคมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.1.28 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 50 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
3	3	1%	2	56.7	57.0	53.3	55.7
			3	41.7	41.0	45.2	43.0
			4	1.6	2.0	1.5	1.3
		10%	2	70.3	68.8	67.9	66.4
			3	22.5	25.9	27.9	21.8
			4	7.2	5.3	4.2	11.8
		20%	2	63.3	61.6	66.3	54.5
			3	23.6	25.3	24.0	24.0
			4	13.1	13.1	9.7	21.5
		30%	2	60.3	55.3	59.6	51.7
			3	23.2	25.7	26.7	26.2
			4	16.5	19.0	13.7	22.1
		40%	2	61.9	56.9	64.6	58.6
			3	19.9	22.9	23.2	25.1
			4	18.2	20.2	12.2	16.3

จากตารางที่ 4.1.28 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1%, 10%, 30% และ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 3 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 20% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 3 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบเคมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.1.29 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 100 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
3	3	1%	2	56.7	54.3	51.5	54.8
			3	42.2	43.8	47.5	44.6
			4	1.1	1.9	1.0	0.6
		10%	2	73.5	68.8	69.5	70.5
			3	20.2	25.2	25.9	23.6
			4	6.3	6.0	4.6	5.9
		20%	2	69.5	69.1	71.8	64.8
			3	21.5	20.9	21.9	19.2
			4	9.0	10.0	6.3	16.0
		30%	2	65.1	60.4	65.8	50.4
			3	22.6	24.9	23.9	24.9
			4	12.3	14.7	10.3	24.7
		40%	2	68.6	58.6	63.8	44.5
			3	19.4	24.8	25.3	32.4
			4	12.0	16.6	10.9	23.1

จากตารางที่ 4.1.29 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1%, 10% และ 20% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 3 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 30% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 3 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบเคมีน และ วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 3 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.1.30 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 300 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
3	3	1%	2	58.7	56.9	53.9	56.1
			3	40.9	41.7	45.4	43.7
			4	0.4	1.4	0.7	0.2
		10%	2	73.4	70.6	70.3	70.2
			3	19.0	22.8	24.4	27.3
			4	7.6	6.6	5.3	2.5
		20%	2	73.2	68.7	70.6	71.7
			3	19.2	22.2	22.5	20.9
			4	7.6	9.1	6.9	7.4
		30%	2	69.3	64.4	65.6	68.7
			3	20.0	22.3	24.8	18.2
			4	10.7	13.3	9.6	13.1
		40%	2	74.0	64.5	71.6	55.0
			3	17.4	23.7	22.8	21.5
			4	8.6	11.8	5.6	23.5

จากตารางที่ 4.1.30 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1%, 20% และ 30% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 3 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 10% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 3 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 3 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบเคมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.1.31 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 50 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
4	2	1%	2	26.1	32.6	26.3	26.3
			3	53.3	43.2	49.7	51.7
			4	20.6	24.2	24.0	22.0
		10%	2	58.0	55.0	53.0	56.8
			3	34.0	35.3	37.2	30.6
			4	8.0	9.7	9.8	12.6
		20%	2	59.8	58.0	58.2	60.0
			3	29.3	29.8	30.9	20.4
			4	10.9	12.2	10.9	19.6
		30%	2	54.6	51.4	52.5	49.0
			3	30.6	31.2	31.9	25.9
			4	14.8	17.4	15.6	25.1
		40%	2	58.0	49.5	52.5	47.5
			3	24.4	30.0	31.2	28.3
			4	17.6	20.5	16.3	24.2

จากตารางที่ 4.1.31 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 4 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบเคมีนเป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 10%, 20%, 30% และ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 4 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.1.32 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 100 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
4	2	1%	2	27.4	34.5	28.7	27.6
			3	52.9	42.8	48.1	49.7
			4	19.7	22.7	23.2	22.7
		10%	2	58.9	54.4	53.4	61.7
			3	32.9	35.7	37.8	27.1
			4	8.2	9.9	8.8	11.2
		20%	2	61.7	57.4	59.9	60.5
			3	26.8	30.5	31.1	24.5
			4	11.5	12.1	9.0	15.0
		30%	2	60.4	56.6	57.8	53.6
			3	26.7	29.0	31.2	24.7
			4	12.9	14.4	11.0	21.7
		40%	2	64.3	53.9	55.7	47.3
			3	25.4	30.1	31.6	26.7
			4	10.3	16.0	12.7	26.0

จากตารางที่ 4.1.32 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลป็น 4 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 10%, 20%, 30% และ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลป็น 4 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.1.33 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 300 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
4	2	1%	2	27.2	32.7	27.5	28.1
			3	53.3	44.6	48.5	50.5
			4	19.5	22.7	24.0	21.4
		10%	2	60.0	55.1	53.2	57.4
			3	32.5	34.9	37.3	30.2
			4	7.5	10.0	9.5	12.4
		20%	2	67.7	63.1	64.3	62.5
			3	24.8	27.9	27.6	23.3
			4	7.5	9.0	8.1	14.2
		30%	2	62.3	56.4	58.7	61.3
			3	28.5	31.7	32.8	20.5
			4	9.2	11.9	8.5	18.2
		40%	2	66.6	51.6	54.7	55.5
			3	23.0	32.9	35.2	25.5
			4	10.4	15.5	10.1	19.0

จากตารางที่ 4.1.33 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 4 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 10%, 20%, 30% และ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 4 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.1.34 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 50 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
4	3	1%	2	25.1	31.2	24.6	25.1
			3	46.3	39.7	41.8	44.6
			4	28.6	29.1	33.6	30.3
		10%	2	62.0	57.0	55.1	58.5
			3	26.1	31.1	33.1	26.3
			4	11.9	11.9	11.8	15.2
		20%	2	63.4	58.1	60.1	55.0
			3	21.1	24.5	25.9	23.7
			4	15.5	17.4	14.0	21.3
		30%	2	59.0	54.9	58.4	52.6
			3	22.8	27.7	27.8	26.7
			4	18.2	17.4	13.8	20.7
		40%	2	61.6	55.6	62.6	56.8
			3	21.0	24.4	24.6	26.0
			4	17.4	20.0	12.8	17.2

จากตารางที่ 4.1.34 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 4 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 10%, 20% และ 30% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 4 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 4 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบเคมีนเป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.1.35 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 100 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
4	3	1%	2	25.3	30.0	23.6	23.7
			3	46.6	37.4	41.4	44.8
			4	28.1	32.6	35.0	31.5
		10%	2	60.2	53.7	53.2	59.3
			3	29.2	34.2	34.7	29.8
			4	10.6	12.1	12.1	10.9
		20%	2	64.1	58.1	62.0	60.6
			3	25.0	27.9	27.4	21.0
			4	10.9	14.0	10.6	18.4
		30%	2	62.7	55.7	61.3	49.9
			3	24.7	28.1	27.2	24.5
			4	12.6	16.2	11.5	25.6
		40%	2	66.0	57.5	63.3	42.9
			3	21.8	25.2	25.2	31.9
			4	12.2	17.3	10.4	25.2

จากตารางที่ 4.1.35 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 4 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 10% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 4 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบเคมีน และ วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 20%, 30% และ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 4 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.1.36 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 300 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
4	3	1%	2	25.9	29.5	24.0	24.6
			3	49.8	42.3	45.1	47.4
			4	24.3	28.2	30.9	28.0
		10%	2	65.9	56.0	55.7	63.0
			3	25.2	31.5	33.2	26.9
			4	8.9	12.5	11.1	10.1
		20%	2	68.8	61.0	62.7	67.4
			3	22.9	26.3	27.5	23.2
			4	8.3	12.7	9.8	9.4
		30%	2	67.8	60.0	65.8	66.2
			3	22.0	24.6	24.7	21.5
			4	10.2	15.4	9.5	12.3
		40%	2	75.1	63.4	71.2	58.3
			3	17.2	21.7	21.0	21.4
			4	7.7	14.9	7.8	20.3

จากตารางที่ 4.1.36 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 4 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 10%, 20% และ 30% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 4 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบเคมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 4 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

4.2 กรณีที่ฐานข้อมูลอยู่ในรูปแบบวงกลม (Spherical or Isotopic)

4.2.1 วิธี Calinski and Harabasz index (Pseudo F)

ตารางที่ 4.2.1 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 50 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
2	2	1%	2	95.6	96.7	96.8	96.0
			3	1.3	1.7	1.5	1.7
			4	3.1	1.6	1.7	2.3
		10%	2	87.4	93.8	93.4	88.2
			3	6.5	4.7	5.1	7.2
			4	6.1	1.5	1.5	4.6
		20%	2	74.7	88.3	89.0	77.8
			3	12.5	7.0	6.9	11.1
			4	12.8	4.7	4.1	11.1
		30%	2	57.2	69.2	70.0	64.8
			3	18.5	15.3	13.9	15.4
			4	24.3	15.5	16.1	19.8
		40%	2	38.6	43.8	43.9	45.5
			3	23.0	24.1	25.8	23.2
			4	38.4	32.1	30.3	31.3

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จากตารางที่ 4.2.1 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1% และ 10% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 2 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบเคมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 20% และ 30% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 2 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 2 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 4.2.2 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 100 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
2	2	1%	2	98.1	99.0	99.0	98.5
			3	0.5	0.9	0.8	0.7
			4	1.4	0.1	0.2	0.8
		10%	2	91.5	95.5	95.9	91.9
			3	5.0	3.3	3.2	4.3
			4	3.5	1.2	0.9	3.8
		20%	2	84.0	93.1	93.7	86.4
			3	9.3	5.6	5.3	8.7
			4	6.7	1.3	1.0	4.9
		30%	2	65.9	83.1	85.2	78.5
			3	16.3	9.3	8.3	10.5
			4	17.8	7.6	6.5	11.0
		40%	2	48.3	55.8	56.9	60.3
			3	21.7	20.7	19.7	16.0
			4	30.0	23.5	23.4	23.7

จากตารางที่ 4.2.2 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 2 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบเคมีนและ วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 10%, 20% และ 30% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 2 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 2 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.2.3 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 300 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
2	2	1%	2	99.3	99.9	99.8	97.6
			3	0.2	0.1	0.2	1.6
			4	0.5	0	0	0.8
		10%	2	94.0	97.4	97.6	91.8
			3	3.1	2.2	2.1	5.7
			4	2.9	0.4	0.3	2.5
		20%	2	89.5	95.5	97.1	90.5
			3	7.3	3.2	2.4	6.5
			4	3.2	1.3	0.5	3.0
		30%	2	75.9	92.6	94.8	87.5
			3	14.6	5.8	3.7	7.8
			4	9.5	1.6	1.5	4.7
		40%	2	56.8	74.4	75.3	78.7
			3	21.1	12.0	11.3	10.1
			4	22.1	13.6	13.4	11.2

จากตารางที่ 4.2.3 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 2 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบเคมีนเป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 10%, 20% และ 30% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 2 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 2 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.2.4 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 50 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
2	3	1%	2	98.7	99.5	99.4	98.5
			3	0.4	0.2	0.2	0.7
			4	0.9	0.3	0.4	0.8
		10%	2	94.1	97.0	97.2	94.4
			3	3.1	1.9	2.1	2.5
			4	2.8	1.1	0.7	3.1
		20%	2	88.2	94.6	96.3	86.7
			3	6.8	3.2	2.7	7.5
			4	5.0	2.2	1.0	5.8
		30%	2	77.0	88.4	91.3	78.6
			3	11.2	7.7	6.7	11.3
			4	11.8	3.9	20	10.1
		40%	2	62.0	72.5	77.5	58.9
			3	19.1	13.4	12.6	19.1
			4	18.9	14.1	9.9	22.0

จากตารางที่ 4.2.4 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 2 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบเคมีนเป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 10%, 20%, 30% และ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 2 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.2.5 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 100 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
2	3	1%	2	99.3	99.7	99.8	99.2
			3	0.4	0	0	0.4
			4	0.3	0.3	0.2	0.4
		10%	2	96.8	98.6	98.9	97.0
			3	1.8	0.9	0.9	2.1
			4	1.4	0.5	0.2	0.9
		20%	2	93.1	98.0	98.9	95.0
			3	4.7	1.2	0.9	2.9
			4	2.2	0.8	0.2	2.1
		30%	2	84.7	93.9	95.0	86.7
			3	8.7	3.0	3.4	7.4
			4	6.6	3.1	1.6	5.9
		40%	2	69.7	87.2	89.6	73.3
			3	15.5	8.4	6.9	11.2
			4	14.8	4.4	3.5	15.5

จากตารางที่ 4.2.5 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1%, 10%, 20%, 30% และ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 2 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบฟัซซีซีมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.2.6 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 300 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
2	3	1%	2	99.5	99.8	99.9	99.1
			3	0.1	0	0	0.6
			4	0.4	0.2	0.1	0.3
		10%	2	97.0	99.8	100.0	97.8
			3	1.4	0	0	1.3
			4	1.6	0.2	0	0.9
		20%	2	95.2	98.4	99.4	95.6
			3	2.9	1.0	0.3	3.0
			4	1.9	0.6	0.3	1.4
		30%	2	92.4	98.4	99.3	94.0
			3	4.9	0.8	0.5	3.2
			4	2.7	0.8	0.2	2.8
		40%	2	81.9	96.3	96.7	88.3
			3	9.9	1.9	1.8	6.8
			4	8.2	1.8	1.5	4.9

จากตารางที่ 4.2.6 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1%, 10%, 20%, 30% และ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 2 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.2.7 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 50 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
3	2	1%	2	6.8	10.9	8.7	7.6
			3	90.7	78.3	86.3	90.8
			4	2.5	10.8	5.0	1.6
		10%	2	54.2	52.4	51.8	42.7
			3	37.8	42.4	43.7	48.9
			4	8.0	5.2	4.5	8.4
		20%	2	60.1	63.3	63.3	53.1
			3	23.8	28.1	29.3	33.2
			4	16.1	8.6	7.4	13.7
		30%	2	50.4	58.4	58.6	45.6
			3	24.2	26.0	26.1	30.6
			4	25.4	15.6	15.3	23.8
		40%	2	40.7	45.5	46.8	39.6
			3	27.7	29.1	31.7	26.8
			4	31.6	25.4	21.5	33.6

จากตารางที่ 4.2.7 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1%, 10%, 20% และ 30% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 3 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 3 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบฟัซซีซีมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.2.8 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 100 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
3	2	1%	2	5.6	8.5	7.6	5.8
			3	92.9	83.6	88.0	92.2
			4	1.5	7.9	4.4	2.0
		10%	2	55.4	54.1	54.0	42.8
			3	39.2	43.8	44.5	54.2
			4	5.4	2.1	1.5	3.0
		20%	2	60.9	62.8	62.9	47.8
			3	27.0	31.2	31.7	41.1
			4	12.1	6.0	5.4	11.1
		30%	2	53.3	60.6	60.3	47.5
			3	24.8	28.4	29.5	32.4
			4	21.9	11.0	10.2	20.1
		40%	2	44.6	52.6	53.2	45.4
			3	26.0	28.9	30.9	29.7
			4	29.4	18.5	15.9	24.9

จากตารางที่ 4.2.8 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 3 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบลำดับชั้น เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 10%, 20% และ 30% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 3 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 3 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบฟัซซี่ ซีมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.2.9 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 300 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
3	2	1%	2	5.9	8.6	7.3	5.4
			3	93.2	81.9	89.0	92.0
			4	0.9	9.5	3.7	2.6
		10%	2	57.7	55.5	55.4	44.9
			3	39.4	43.5	44.0	52.0
			4	2.9	1.0	0.6	3.1
		20%	2	65.5	67.7	67.7	49.7
			3	26.9	29.0	29.1	42.8
			4	7.6	3.3	3.2	7.5
		30%	2	59.0	63.1	64.0	45.5
			3	24.8	29.9	31.1	41.1
			4	16.2	7.0	4.9	13.4
		40%	2	48.0	52.7	53.8	38.4
			3	29.3	34.5	37.5	38.4
			4	22.7	12.8	8.7	23.2

จากตารางที่ 4.2.9 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 3 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบลำดับชั้น เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 10%, 20%, 30% และ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 3 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.2.10 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 50 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
3	3	1%	2	15.0	19.5	14.5	11.0
			3	84.4	75.9	85.2	88.6
			4	0.6	4.6	0.3	0.4
		10%	2	65.5	64.5	63.8	47.7
			3	31.5	34.3	36.0	48.7
			4	3.0	1.2	0.2	3.6
		20%	2	68.3	73.5	74.5	54.3
			3	22.5	23.8	23.7	36.9
			4	9.2	2.7	1.8	8.8
		30%	2	62.6	73.3	76.4	46.5
			3	22.0	19.5	18.7	32.7
			4	15.4	7.2	4.9	20.8
		40%	2	53.7	64.9	68.7	45.0
			3	23.6	21.0	22.0	26.4
			4	22.7	14.1	9.3	28.6

จากตารางที่ 4.2.10 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1%, 10%, 20%, 30% และ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 3 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.2.11 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 100 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
3	3	1%	2	12.6	18.2	13.1	9.8
			3	87.2	78.5	86.7	90.0
			4	0.2	3.3	0.2	0.2
		10%	2	64.1	63.9	63.7	45.7
			3	33.7	35.4	35.7	52.4
			4	2.2	0.7	0.6	1.9
		20%	2	71.5	74.3	74.7	74.7
			3	21.9	23.3	23.3	24.7
			4	6.6	2.4	2.4	0.6
		30%	2	63.8	75.2	77.4	48.4
			3	25.0	20.1	20.0	36.5
			4	11.2	4.7	2.6	15.1
		40%	2	58.8	72.4	76.4	44.6
			3	22.0	17.7	17.1	29.9
			4	19.2	9.9	6.5	25.5

จากตารางที่ 4.2.11 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1%, 10%, 20%, 30% และ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 3 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.2.12 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 300 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
3	3	1%	2	13.6	18.2	13.5	9.5
			3	86.3	79.6	86.5	89.3
			4	0.1	2.2	0	1.2
		10%	2	64.4	62.7	61.4	42.1
			3	35.2	37.0	38.6	56.8
			4	0.4	0.3	0	1.1
		20%	2	69.5	74.9	74.3	43.8
			3	24.7	24.1	25.5	52.2
			4	5.8	1.0	0.2	4.0
		30%	2	71.1	79.5	81.9	44.6
			3	21.6	17.4	17.0	43.4
			4	7.3	3.1	1.1	12.0
		40%	2	64.5	81.3	82.9	45.8
			3	22.7	14.7	14.6	38.4
			4	12.8	4.0	2.5	15.8

จากตารางที่ 4.2.12 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1%, 10%, 20%, 30% และ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 3 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.2.13 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 50 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
4	2	1%	2	0.4	7.1	3.2	0.8
			3	5.4	27.7	19.5	9.8
			4	94.2	65.2	77.3	89.4
		10%	2	28.7	28.7	27.1	23.9
			3	46.7	47.8	49.2	43.8
			4	24.6	23.5	23.7	32.3
		20%	2	45.9	46.8	46.9	36.7
			3	32.7	37.5	37.1	37.9
			4	21.4	15.7	16.0	25.4
		30%	2	42.5	46.6	47.4	36.9
			3	32.5	36.3	36.7	33.7
			4	25.0	17.1	15.9	29.4
		40%	2	36.7	37.6	39.3	32.7
			3	28.7	35.8	36.6	30.2
			4	34.6	26.6	24.1	37.1

จากตารางที่ 4.2.13 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 4 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบลำดับชั้น เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 10%, 20%, 30% และ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 4 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.2.14 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 100 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
4	2	1%	2	0.3	6.8	2.5	0.7
			3	5.9	28.0	18.2	7.1
			4	93.8	65.2	79.3	92.2
		10%	2	29.0	24.8	23.9	22.2
			3	47.1	51.5	54.0	46.3
			4	23.9	23.7	22.1	31.5
		20%	2	45.4	46.4	46.5	33.8
			3	35.4	39.7	39.9	40.0
			4	19.2	13.9	13.6	26.2
		30%	2	43.0	47.5	48.7	35.6
			3	31.6	38.3	39.2	36.7
			4	25.4	14.2	12.1	27.7
		40%	2	37.7	38.1	39.1	27.6
			3	31.0	36.2	38.7	33.1
			4	31.3	25.7	22.2	39.3

จากตารางที่ 4.2.14 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 4 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบลำดับชั้น เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 10%, 20%, 30% และ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 4 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.2.15 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 300 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
4	2	1%	2	0.4	7.4	2.1	0.5
			3	6.3	26.7	20.7	7.0
			4	93.3	65.9	77.2	92.5
		10%	2	30.4	28.7	27.9	24.0
			3	46.8	46.0	48.8	36.5
			4	22.8	25.3	23.3	39.5
		20%	2	49.5	47.7	47.6	35.5
			3	34.7	40.7	41.3	34.2
			4	15.8	11.6	11.1	30.3
		30%	2	45.7	47.6	47.9	30.2
			3	33.2	41.8	41.9	38.0
			4	21.1	10.6	10.2	31.8
		40%	2	40.0	41.6	42.6	27.7
			3	34.5	42.2	43.2	33.6
			4	25.5	16.2	14.2	38.7

จากตารางที่ 4.2.15 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 4 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบลำดับชั้น เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 10%, 20%, 30% และ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 4 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.2.16 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 50 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
4	3	1%	2	4.1	10.0	4.1	3.2
			3	14.3	25.6	17.0	10.9
			4	81.6	64.4	78.9	85.9
		10%	2	45.7	44.0	42.7	29.0
			3	38.0	41.8	43.0	40.4
			4	16.3	14.2	14.3	30.6
		20%	2	55.8	60.2	60.1	35.4
			3	30.7	32.4	34.1	37.1
			4	13.5	7.4	5.8	27.5
		30%	2	54.9	64.5	66.3	36.1
			3	26.7	25.7	26.0	34.1
			4	18.4	9.8	7.7	29.8
		40%	2	46.5	56.3	61.8	28.4
			3	27.6	27.3	26.1	36.4
			4	25.9	16.4	12.1	35.2

จากตารางที่ 4.2.16 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1%, 10%, 20%, 30% และ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 4 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.2.17 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 100 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
4	3	1%	2	4.1	10.4	4.2	2.9
			3	13.3	25.5	16.9	8.4
			4	82.6	64.1	78.9	88.7
		10%	2	44.4	41.5	40.1	28.9
			3	41.3	45.5	46.7	35.0
			4	14.3	13.0	13.2	36.1
		20%	2	60.1	63.6	61.8	25.4
			3	29.5	30.7	32.3	39.9
			4	10.4	5.7	5.9	25.4
		30%	2	58.6	66.7	68.1	33.5
			3	26.4	25.6	26.8	37.3
			4	15.0	7.7	5.1	29.2
		40%	2	54.6	66.4	70.2	31.2
			3	26.7	23.0	23.2	34.7
			4	18.7	10.6	6.6	34.1

จากตารางที่ 4.2.17 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1%, 10%, 20%, 30% และ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 4 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.2.18 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Pseudo F ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 300 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
4	3	1%	2	3.3	7.7	3.8	2.6
			3	15.3	25.5	18.5	11.2
			4	81.4	66.8	77.7	86.2
		10%	2	46.5	44.3	42.9	29.8
			3	39.7	43.4	44.6	32.7
			4	13.8	12.3	12.5	37.5
		20%	2	59.8	61.5	60.8	30.7
			3	29.9	32.6	34.1	35.8
			4	10.3	5.9	5.1	33.5
		30%	2	59.9	70.3	71.4	28.3
			3	28.2	25.5	25.2	38.2
			4	11.9	4.2	3.4	33.6
		40%	2	56.7	71.0	75.1	28.2
			3	26.5	20.8	19.9	38.0
			4	16.8	8.2	5.0	33.8

จากตารางที่ 4.2.18 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1%, 10%, 20%, 30% และ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 4 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

4.2.2 วิธี Silhouette width

ตารางที่ 4.2.19 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 50 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
2	2	1%	2	97.5	97.7	97.6	97.8
			3	0.8	1.2	1.1	0.9
			4	1.7	1.1	1.3	1.3
		10%	2	88.3	92.6	93.7	87.8
			3	7.7	5.7	5.4	8.3
			4	4.0	1.7	0.9	3.9
		20%	2	84.1	90.7	92.4	76.9
			3	11.2	7.8	6.6	13.9
			4	4.7	1.5	1.0	9.2
		30%	2	72.2	82.5	85.5	67.7
			3	18.6	13.1	9.9	21.0
			4	9.2	4.4	4.6	11.3
		40%	2	60.2	65.0	66.5	63.9
			3	21.1	21.3	20.5	23.5
			4	18.7	13.7	13.0	12.6

จากตารางที่ 4.2.19 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 2 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 10%, 20%, 30% และ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 2 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.2.20 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 100 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
2	2	1%	2	98.5	99.1	99.1	98.2
			3	0.7	0.5	0.6	1.2
			4	0.8	0.4	0.3	0.6
		10%	2	92.2	93.9	95.1	91.7
			3	4.9	4.6	4.2	5.1
			4	2.9	1.5	0.7	3.2
		20%	2	87.7	91.6	95.1	82.4
			3	9.8	7.6	4.6	12.4
			4	2.5	0.8	0.3	5.2
		30%	2	82.4	90.4	95.3	78.1
			3	12.7	8.1	4.2	13.5
			4	4.9	1.5	0.5	8.4
		40%	2	69.5	82.3	83.8	68.4
			3	19.3	12.4	10.6	19.6
			4	11.2	5.3	5.6	12.0

จากตารางที่ 4.2.20 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 2 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบเคมีน และ วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 10%, 20%, 30% และ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 2 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.2.21 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 300 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
2	2	1%	2	99.1	99.2	99.2	97.5
			3	0.3	0.5	0.6	1.8
			4	0.6	0.3	0.2	0.7
		10%	2	93.6	94.1	95.2	92.1
			3	4.4	5.2	4.5	6.1
			4	2.0	0.7	0.3	1.8
		20%	2	90.6	92.5	97.2	89.9
			3	8.4	6.4	2.6	7.9
			4	1.0	1.1	0.2	2.2
		30%	2	86.5	95.1	98.4	84.4
			3	11.0	4.3	1.3	11.5
			4	2.5	0.6	0.3	4.1
		40%	2	80.8	95.4	96.6	79.4
			3	14.7	3.1	2.8	15.1
			4	4.5	1.5	0.6	5.5

จากตารางที่ 4.2.21 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 2 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบเคมีน และ วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 10%, 20%, 30% และ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 2 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.2.22 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 50 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
2	3	1%	2	97.1	97.0	97.3	97.2
			3	0.6	1.1	1.1	1.0
			4	2.3	1.9	1.6	1.8
		10%	2	88.1	92.0	95.4	91.4
			3	6.3	5.7	4.1	4.7
			4	5.6	2.3	0.5	3.9
		20%	2	83.4	91.2	97.0	79.0
			3	11.1	6.3	2.8	12.2
			4	5.5	2.5	2	8.8
		30%	2	78.3	87.4	92.9	73.2
			3	12.2	7.3	5.1	15.7
			4	9.5	5.3	2.0	11.1
		40%	2	67.1	71.6	76.9	66.2
			3	16.7	12.9	11.1	18.4
			4	16.2	15.5	12.0	15.4

จากตารางที่ 4.2.22 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1%, 10%, 20%, 30% และ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 2 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบฟัซซีซีมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.2.23 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 100 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
2	3	1%	2	97.2	97.1	97.5	98.1
			3	0.6	1.1	1.1	0.8
			4	2.2	1.8	1.4	1.1
		10%	2	91.5	94.6	97.4	94.7
			3	4.7	2.8	2.5	4.0
			4	3.8	2.6	0.1	1.3
		20%	2	89.2	95.1	99.2	90.1
			3	7.7	3.7	0.8	6.0
			4	3.1	1.2	0	3.9
		30%	2	87.0	94.2	98.5	81.2
			3	9.1	3.5	1.2	13.2
			4	3.9	2.3	0.3	5.6
		40%	2	74.5	84.9	90.1	71.8
			3	14.4	7.2	4.2	16.0
			4	11.1	7.9	5.7	12.2

จากตารางที่ 4.2.23 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 2 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 10%, 20%, 30% และ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 2 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.2.24 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 300 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
2	3	1%	2	97.3	96.9	97.3	97.9
			3	0.4	0.4	0.6	1.2
			4	2.3	2.7	2.1	0.9
		10%	2	91.0	94.6	99.3	96.1
			3	3.1	1.7	0.7	3.3
			4	5.9	3.7	0	0.6
		20%	2	92.6	97.5	100.0	94.5
			3	4.5	1.5	0	4.4
			4	2.9	1.0	0	1.1
		30%	2	91.3	98.6	99.6	93.1
			3	6.1	0.5	0.2	5.4
			4	2.6	0.9	0.2	1.5
		40%	2	86.6	97.2	98.7	85.7
			3	8.3	1.1	0.3	9.6
			4	5.1	1.7	1.0	4.7

จากตารางที่ 4.2.24 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 2 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 10%, 20%, 30% และ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 2 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.2.25 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 50 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
3	2	1%	2	61.3	60.8	57.6	57.2
			3	38.3	38.0	42.0	42.2
			4	0.4	1.2	0.4	0.6
		10%	2	77.0	73.7	73.7	65.4
			3	18.5	22.1	23.7	27.1
			4	4.5	4.2	2.6	7.5
		20%	2	74.1	73.2	73.5	63.7
			3	16.9	20.1	21.4	24.0
			4	9.0	6.7	5.1	12.3
		30%	2	65.8	66.2	70.0	58.5
			3	21.6	22.9	22.1	24.6
			4	12.6	10.9	7.9	16.9
		40%	2	61.4	61.1	63.7	57.2
			3	24.6	26.3	27.5	23.5
			4	14.0	12.6	8.8	19.3

จากตารางที่ 4.2.25 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1%, 10%, 20% และ 30%โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 3 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 3 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบฟัซซีซีมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.2.26 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 100 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
3	2	1%	2	59.9	58.0	56.0	56.2
			3	39.9	41.1	43.8	43.3
			4	0.2	0.9	0.2	0.5
		10%	2	77.6	74.6	74.2	67.7
			3	19.1	23.2	24.7	28.1
			4	3.3	2.2	1.1	4.2
		20%	2	74.2	71.0	72.0	63.0
			3	18.7	22.8	23.7	26.2
			4	7.1	6.2	4.3	10.8
		30%	2	67.9	67.7	70.3	58.7
			3	21.3	23.3	23.4	25.2
			4	10.8	9.0	6.3	16.1
		40%	2	66.7	67.8	71.8	60.3
			3	22.1	23.3	23.0	24.7
			4	11.2	8.9	5.2	15.0

จากตารางที่ 4.2.26 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 3 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 10%, 20%, 30% และ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 3 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.2.27 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 300 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
3	2	1%	2	58.9	57.2	55.2	54.6
			3	40.6	41.7	44.3	43.9
			4	0.5	1.1	0.5	1.5
		10%	2	80.4	75.7	75.1	69.0
			3	18.0	22.2	24.0	26.9
			4	1.6	2.1	0.9	4.1
		20%	2	79.2	75.5	75.8	65.7
			3	15.9	20.2	21.3	27.5
			4	4.9	4.3	2.9	6.8
		30%	2	74.4	69.7	72.6	60.0
			3	18.0	23.7	23.8	27.0
			4	7.6	6.6	3.6	13.0
		40%	2	71.3	68.3	74.0	55.3
			3	21.5	22.3	22.1	26.7
			4	7.2	9.4	3.9	18.0

จากตารางที่ 4.2.27 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 3 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 10%, 20%, 30% และ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 3 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.2.28 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 50 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
3	3	1%	2	54.7	54.7	49.9	47.3
			3	44.2	43.6	49.2	52.2
			4	1.1	1.7	0.9	0.5
		10%	2	70.5	69.3	67.0	59.1
			3	22.0	25.9	30.8	34.5
			4	7.5	4.8	2.2	6.4
		20%	2	68.8	69.6	73.9	59.5
			3	17.1	20.8	20.4	28.9
			4	14.1	9.6	5.7	11.6
		30%	2	62.9	68.2	78.9	52.7
			3	20.8	19.7	16.2	26.4
			4	16.3	12.1	4.9	20.9
		40%	2	63.2	66.8	74.8	55.3
			3	19.5	14.8	15.6	21.5
			4	17.3	18.4	9.6	23.2

จากตารางที่ 4.2.28 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1%, 10%, 20%, 30% และ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 3 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.2.29 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 100 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
3	3	1%	2	53.8	52.8	49.3	48.1
			3	45.6	45.9	50.1	51.5
			4	0.6	1.3	0.6	0.4
		10%	2	74.6	69.0	67.7	58.7
			3	19.4	26.3	30.9	36.2
			4	6.0	4.7	1.4	5.1
		20%	2	70.7	69.2	73.5	58.1
			3	17.0	21.3	23.7	30.7
			4	12.3	9.5	2.8	11.2
		30%	2	67.4	68.9	79.1	58.1
			3	17.8	20.1	17.2	24.7
			4	14.8	11.0	3.7	17.2
		40%	2	65.5	71.4	80.4	57.7
			3	18.4	15.0	13.5	24.1
			4	16.1	13.6	6.1	18.2

จากตารางที่ 4.2.29 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1%, 10%, 20%, 30% และ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 3 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.2.30 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 300 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
3	3	1%	2	54.1	51.6	49.4	47.0
			3	45.5	47.2	50.2	51.7
			4	0.4	1.2	0.4	1.3
		10%	2	74.2	66.9	67.6	55.5
			3	21.9	30.0	31.9	42.1
			4	3.9	3.1	0.5	2.4
		20%	2	70.8	69.1	71.5	53.6
			3	17.4	20.7	25.0	39.3
			4	11.8	10.2	3.5	7.1
		30%	2	70.6	71.6	84.6	56.3
			3	18.5	17.2	13.7	31.8
			4	10.9	11.2	1.7	11.9
		40%	2	74.1	81.4	91.2	58.3
			3	17.6	11.7	7.0	26.6
			4	8.3	6.9	1.8	15.1

จากตารางที่ 4.2.30 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1%, 10%, 20%, 30% และ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 3 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.2.31 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 50 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
4	2	1%	2	26.6	34.5	27.8	25.5
			3	52.6	40.9	47.8	49.8
			4	20.8	24.6	24.4	24.7
		10%	2	62.1	58.1	56.4	51.9
			3	32.1	35.5	37.4	34.6
			4	5.8	6.4	6.2	13.5
		20%	2	67.5	64.1	64.8	55.1
			3	23.0	26.9	27.9	27.2
			4	9.5	9.0	7.3	17.7
		30%	2	62.3	60.6	61.9	54.1
			3	26.0	27.9	30.3	26.4
			4	11.7	11.5	7.8	19.5
		40%	2	58.8	54.7	58.2	48.2
			3	26.2	28.8	30.0	25.1
			4	15.0	16.5	11.8	26.7

จากตารางที่ 4.2.31 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1%, 10%, 20%, 30% และ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 4 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.2.32 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 100 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
4	2	1%	2	26.6	32.6	28.1	25.5
			3	50.9	42.2	45.0	46.8
			4	22.5	25.2	26.9	27.7
		10%	2	56.3	56.3	54.5	50.0
			3	35.2	36.4	38.6	35.0
			4	6.5	7.3	6.9	15.0
		20%	2	69.0	64.0	63.3	55.3
			3	23.2	27.4	29.3	26.3
			4	7.8	8.6	7.4	18.4
		30%	2	63.2	60.4	61.4	51.6
			3	25.5	30.1	32.0	25.5
			4	11.3	9.5	6.6	22.9
		40%	2	41.0	38.0	37.8	48.3
			3	38.9	39.7	42.4	27.6
			4	20.1	22.3	19.8	24.1

จากตารางที่ 4.2.32 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1%, 10%, 20%, 30% และ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 4 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.2.33 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม ขนาดตัวอย่างกลุ่มละ 300 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
4	2	1%	2	28.7	36.1	30.1	28.1
			3	51.4	39.4	45.8	47.1
			4	19.9	24.5	24.1	24.8
		10%	2	59.0	54.0	52.0	49.2
			3	35.4	38.2	40.1	32.7
			4	5.6	7.8	7.9	18.1
		20%	2	70.7	62.9	64.6	55.5
			3	23.1	29.7	29.2	24.9
			4	6.2	7.4	6.2	19.6
		30%	2	67.5	59.5	61.6	48.6
			3	25.0	31.2	32.6	26.6
			4	7.5	9.3	5.8	24.8
		40%	2	61.7	54.3	58.0	46.2
			3	29.4	32.7	33.7	25.3
			4	8.9	13.0	8.3	28.5

จากตารางที่ 4.2.33 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1%, 10%, 20%, 30% และ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 4 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.2.34 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 50 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
4	3	1%	2	26.1	30.8	25.8	23.8
			3	47.6	41.3	43.6	43.0
			4	26.3	27.9	30.6	33.2
		10%	2	58.3	51.7	51.3	44.4
			3	29.1	32.6	35.1	30.9
			4	12.6	15.7	13.6	24.7
		20%	2	61.3	57.1	60.6	47.1
			3	24.3	27.4	30.2	25.7
			4	14.1	15.5	9.2	27.2
		30%	2	59.4	57.7	66.9	48.1
			3	21.7	25.2	22.2	24.9
			4	18.9	17.1	10.9	27.0
		40%	2	57.5	53.4	65.6	46.5
			3	21.7	20.8	20.2	26.3
			4	20.8	25.8	14.2	27.2

จากตารางที่ 4.2.34 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1%, 10%, 20%, 30% และ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 4 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.2.35 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 100 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
4	3	1%	2	23.5	28.9	24.0	21.0
			3	48.3	40.5	42.6	43.2
			4	28.2	30.6	33.4	35.8
		10%	2	57.5	52.4	51.5	42.9
			3	30.7	32.9	35.0	28.6
			4	11.8	14.7	13.5	28.5
		20%	2	67.9	58.3	62.9	46.4
			3	21.5	26.8	27.2	28.3
			4	10.6	14.9	9.9	25.3
		30%	2	61.8	58.8	70.0	46.3
			3	21.2	23.3	21.8	28.3
			4	17.0	17.9	8.2	25.4
		40%	2	65.1	63.5	75.3	51.6
			3	20.8	18.8	16.3	23.2
			4	14.1	17.7	8.4	25.2

จากตารางที่ 4.2.35 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1%, 10%, 20%, 30% และ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 4 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

ตารางที่ 4.2.36 เปอร์เซ็นต์ของจำนวนกลุ่มที่มีค่า Silhouette width ดีที่สุดในแต่ละรอบ สำหรับข้อมูลที่มีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม (ขนาดตัวอย่างกลุ่มละ 300 ตัวอย่าง และจำนวนตัวแปรที่ใช้ในการศึกษา 3 ตัวแปร)

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
4	3	1%	2	25.1	29.4	24.0	23.7
			3	47.6	38.3	42.7	41.8
			4	27.3	32.3	33.3	34.5
		10%	2	60.1	53.1	50.8	42.7
			3	29.5	32.6	34.6	26.8
			4	10.4	14.3	14.6	30.5
		20%	2	67.4	58.3	58.9	42.2
			3	20.6	27.1	32.4	25.2
			4	12.0	14.6	8.7	32.6
		30%	2	67.9	62.2	72.9	42.3
			3	19.9	19.7	20.3	28.0
			4	12.2	18.1	6.8	29.7
		40%	2	69.3	65.8	82.9	49.9
			3	18.8	16.3	12.8	26.3
			4	11.9	17.9	4.3	23.8

จากตารางที่ 4.2.36 พบว่า

- กรณีที่ค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 1%, 10%, 20%, 30% และ 40% โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม เมื่อทำการจัดกลุ่มข้อมูลเป็น 4 กลุ่ม ตามจำนวนกลุ่มของการจำลอง วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพดีที่สุด

4.3 ข้อมูลจริง

ข้อมูลรูปพรรณสัณฐานของปู 200 หน่วยตัวอย่าง โดยเก็บข้อมูลตามขนาดรูปพรรณสัณฐาน 5 ประเภท คือ หน้าผากปู (frontal lobe: FL) หน่วยเซนติเมตร, ความกว้างด้านหลัง (rear width: RW) หน่วยเซนติเมตร, ความยาวกระดอง (carapace length: CL) หน่วยเซนติเมตร, ความกว้างกระดอง (carapace width: CW) หน่วยเซนติเมตร, ความสูงของตัวปู (body depth: BD) หน่วยเซนติเมตร ซึ่งตัวอย่างข้อมูลแสดงดังตาราง

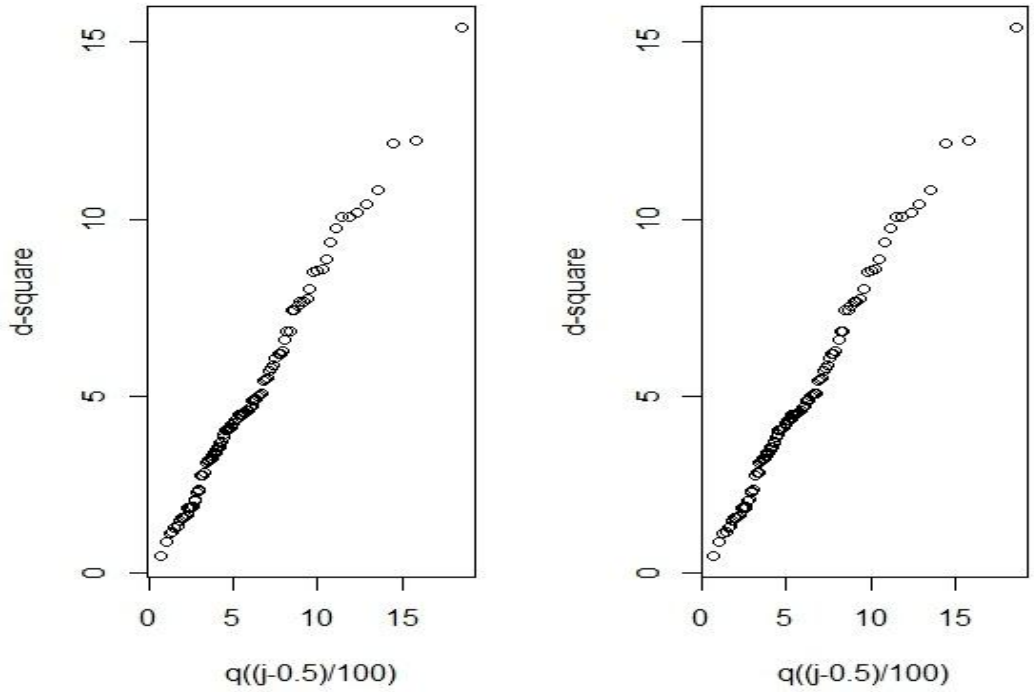
ตาราง แสดงข้อมูลรูปพรรณสัณฐานของปู 200 หน่วยตัวอย่าง

หน่วยตัวอย่าง	FL	RW	CL	CW	BD
1	0.81	0.67	1.61	1.90	0.70
2	0.88	0.77	1.81	2.08	0.74
3	0.92	0.78	1.90	2.24	0.77
4	0.96	0.79	2.01	2.31	0.82
5	0.98	0.80	2.03	2.30	0.82
6	1.08	0.90	2.30	2.65	0.98
7	1.11	0.99	2.38	2.71	0.98
8	1.16	0.91	2.45	2.83	1.04
⋮	⋮	⋮	⋮	⋮	⋮
199	2.25	1.72	4.30	4.87	1.98
200	2.31	2.02	4.62	5.25	2.11

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

4.3.1 ข้อมูลรูปพรรณสัณฐานของปู 200 หน่วยตัวอย่าง ซึ่งแบ่งเป็น 2 กลุ่ม ตามเพศ คือ เพศผู้ และเพศเมีย อย่างละ 100 ตัว

นำข้อมูลรูปพรรณสัณฐานของปูมาทำการพล็อตกราฟเพื่อทดสอบว่าข้อมูลมีการแจกแจงปกติหรือไม่



ภาพที่ 4.3.1 กราฟ Chi-square ของข้อมูลตามเพศ

(กราฟซ้าย: เพศผู้ กราฟขวา: เพศเมีย)

กราฟ Chi-square Q-Q plot เป็นการตรวจสอบข้อมูลที่มีการแจกแจงหลายตัวแปร โดยใช้ค่าระยะทางกำลังสองของ Mahalanobis กับค่า Chi-square ในการพล็อตซึ่งจากข้อมูลรูปพรรณสัณฐานของปูนำมาทำการพล็อตกราฟ พบว่า ข้อมูลมีการแจกแจงปกติ เนื่องจาก กราฟที่ได้มีแนวโน้มเป็นเชิงเส้นตรง และจากตรวจสอบค่าการซ้อนทับกันของข้อมูล โดยใช้ฟังก์ชัน overlap ในโปรแกรม R พบว่าข้อมูลมีค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 6.482131% จากข้อมูลปูเพศผู้มีเมทริกซ์ความแปรปรวนร่วมเท่ากับ

$$\Sigma_1 = \begin{bmatrix} 0.11994994 & 0.07244283 & 0.2545926 & 0.2785104 & 0.1202317 \\ 0.07244283 & 0.04667778 & 0.1587688 & 0.1761999 & 0.0729704 \\ 0.25459259 & 0.15876879 & 0.5581889 & 0.6190702 & 0.2558304 \\ 0.27851044 & 0.17619990 & 0.6190702 & 0.6939303 & 0.2794377 \\ 0.12023174 & 0.07297040 & 0.2558304 & 0.2794377 & 0.1220963 \end{bmatrix}$$

ข้อมูลปูเพศเมียมีเมทริกซ์ความแปรปรวนร่วมเท่ากับ

$$\Sigma_2 = \begin{bmatrix} 0.12516945 & 0.09382541 & 0.2327271 & 0.2534044 & 0.11647911 \\ 0.09382541 & 0.07511445 & 0.1812493 & 0.1992736 & 0.08918901 \\ 0.23272707 & 0.18124929 & 0.4493010 & 0.4929869 & 0.22140263 \\ 0.25340444 & 0.19927364 & 0.4929869 & 0.5447808 & 0.24141596 \\ 0.11647911 & 0.08918901 & 0.2214026 & 0.2414160 & 0.11177196 \end{bmatrix}$$

เมื่อนำข้อมูลข้างต้นมาทำการจัดกลุ่มข้อมูลออกเป็น 2 กลุ่ม ตามจำนวนการแบ่งกลุ่มตามเพศ ด้วยวิธีการจัดกลุ่มแบบลำดับชั้น, วิธีการจัดกลุ่มแบบเคมีน, วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน และวิธีการจัดกลุ่มแบบอัลกอริทึม EM โดยใช้วิธี Calinski and Harabasz index (Pseudo F) และวิธี Silhouette width เป็นเกณฑ์ในการวัดประสิทธิภาพของวิธีการจัดกลุ่มข้อมูลให้ผลดังนี้

4.3.1.1 วิธี Calinski and Harabasz index (Pseudo F)

ตารางที่ 4.3.1 ค่า Pseudo F โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม ขนาดตัวอย่างกลุ่มละ 100 ตัวอย่าง

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
2	5	6.482131%	2	359.3804	390.9751	390.9375	309.639273
			3	392.3633	433.6940	436.7241	9.657345
			4	439.5284	546.8992	542.1265	113.013699

หมายเหตุ ตัวอักษรสีแดง แทน ค่า Pseudo F ที่ดีที่สุดในแต่ละวิธีการจัดกลุ่ม
ช่องแถบสี แทน วิธีการจัดกลุ่มที่ดีที่สุด เมื่อจำนวนกลุ่มของการจัดกลุ่มข้อมูล เท่ากับจำนวนกลุ่มของการซ้อนทับ

จากตารางที่ 4.3.1

ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม ขนาดตัวอย่างกลุ่มละ 100 ตัวอย่าง ข้อมูลมีค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 6.482131% นำมาทำการจัดกลุ่มข้อมูลออกเป็น 2, 3 และ 4 กลุ่ม ซึ่งในแต่ละวิธีจะพบค่า Pseudo F ที่ดีที่สุด แต่เมื่อพิจารณากรณีที่จำนวนกลุ่มของการจัดกลุ่มข้อมูลเท่ากับจำนวนกลุ่มของการซ้อนทับ พบว่า วิธีการจัดกลุ่มแบบอัลกอริทึม EM มีประสิทธิภาพการจัดกลุ่มดีที่สุด

4.3.1.2 วิธี Silhouette width

ตารางที่ 4.3.2 ค่า Silhouette width โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม ขนาดตัวอย่างกลุ่มละ 100 ตัวอย่าง

จำนวนกลุ่มซ้อนทับ	จำนวนตัวแปร	$\bar{\omega}$	k	วิธีการจัดกลุ่ม			
				HC	KC	FC	EM
2	5	6.482131%	2	0.5374496	0.5455182	0.5445917	0.50044064
			3	0.4297302	0.4846869	0.4748246	-0.05997106
			4	0.4552584	0.4842591	0.4800032	0.06683738

หมายเหตุ ตัวอักษรสีแดง แทน ค่า Pseudo F ที่ดีที่สุดในแต่ละวิธีการจัดกลุ่ม
ช่องแถบสี แทน วิธีการจัดกลุ่มที่ดีที่สุด เมื่อจำนวนกลุ่มของการจัดกลุ่มข้อมูล เท่ากับจำนวนกลุ่มของการซ้อนทับ

จากตารางที่ 4.3.2

ข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม ขนาดตัวอย่างกลุ่มละ 100 ตัวอย่าง ข้อมูลมีค่าการซ้อนทับเฉลี่ย $(\bar{\omega})$ เท่ากับ 6.482131% นำมาทำการจัดกลุ่มข้อมูลออกเป็น 2, 3 และ 4 กลุ่ม ซึ่งในแต่ละวิธีจะพบค่า Silhouette width ที่ดีที่สุด แต่เมื่อพิจารณากรณีที่จำนวนกลุ่มของการจัดกลุ่มข้อมูล เท่ากับจำนวนกลุ่มของการซ้อนทับ พบว่า วิธีการจัดกลุ่มแบบลำดับชั้น, วิธีการจัดกลุ่มแบบเคมีน, วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน และ วิธีการจัดกลุ่มแบบอัลกอริทึม EM มีประสิทธิภาพการจัดกลุ่มดีที่สุดทั้ง 4 วิธี

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทที่ 5

สรุปผลงานวิจัย อภิปรายผล และข้อเสนอแนะ

งานวิจัยนี้มีวัตถุประสงค์เพื่อเพื่อเปรียบเทียบประสิทธิภาพวิธีการจัดกลุ่มสำหรับข้อมูลที่มีการแจกแจงปกติแบบผสม โดยใช้วิธีการจัดกลุ่ม 4 วิธี คือ วิธีการจัดกลุ่มแบบลำดับชั้น, วิธีการจัดกลุ่มแบบเคมีน, วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน และวิธีการจัดกลุ่มแบบอัลกอริทึม EM นำมาศึกษากับข้อมูล 2 ชนิด คือ ข้อมูลทำการจำลองขึ้นมา และข้อมูลจริง โดยใช้วิธี Calinski and Harabasz index (Pseudo F) และวิธี Silhouette width เป็นเกณฑ์ในการวัดประสิทธิภาพของวิธีการจัดกลุ่มข้อมูล

ในการนำเสนอผลงานวิจัยจะแสดงในรูปแบบตาราง มีสัญลักษณ์ที่ใช้ ดังนี้

$\bar{\omega}$	แทน	ค่าการซ้อนทับเฉลี่ย
k	แทน	จำนวนกลุ่มของการจัดกลุ่มข้อมูล
K	แทน	จำนวนกลุ่มของข้อมูลที่มีการซ้อนทับกัน
p	แทน	จำนวนตัวแปรที่ใช้ในการศึกษา
HC	แทน	วิธีการจัดกลุ่มแบบลำดับชั้น ด้วยวิธี Ward
KC	แทน	วิธีการจัดกลุ่มแบบเคมีน
FC	แทน	วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน
EM	แทน	วิธีการจัดกลุ่มแบบอัลกอริทึม EM
NonSph	แทน	กรณีที่ฐานข้อมูลอยู่ในรูปแบบวงรี (Non Spherical)
Sph	แทน	กรณีที่ฐานข้อมูลอยู่ในรูปแบบวงกลม (Spherical or Isotopic)
CH	แทน	วิธี Calinski and Harabasz index (Pseudo F)
Sil	แทน	วิธี Silhouette width

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

5.1 สรุปผลงานวิจัย อภิปรายผล

5.1.1 จำลองข้อมูลกรณีที่ฐานข้อมูลอยู่ในรูปแบบวงรี (Non Spherical) และกรณีที่ฐานข้อมูลอยู่ในรูปแบบวงกลม (Spherical or Isotopic) โดยที่ข้อมูลมีอัตราการซ้อนทับเฉลี่ย $(\bar{\omega})$ ที่ 1%, 10%, 20%, 30% และ 40% จำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม มีจำนวนตัวแปรที่ใช้ศึกษา คือ 2 และ 3 ตัวแปร โดยจำลองข้อมูลขนาดกลุ่มละ 50, 100 และ 300 ตัวอย่าง กำหนดจำนวนกลุ่มของการจัดกลุ่มข้อมูลเป็น 2 กลุ่มตามการจำลอง เพื่อทำการเปรียบเทียบประสิทธิภาพวิธีการจัดกลุ่ม 4 วิธี โดยใช้วิธี Calinski and Harabasz index (Pseudo F) และวิธี Silhouette width เป็นเกณฑ์การวัดประสิทธิภาพของวิธีการจัดกลุ่มข้อมูล สามารถสรุปผลได้ดังนี้

ตารางที่ 5.1.1 สรุปผลวิธีการจัดกลุ่มข้อมูลด้วยค่า Pseudo F และค่า Silhouette width โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม จำนวนตัวแปรที่ใช้ศึกษา 2 ตัวแปร

K	p	k	n	$\bar{\omega}$	วิธีการจัดกลุ่ม			
					NonSph		Sph	
					CH	Sil	CH	Sil
2	2	2	50	1%	EM	EM	KC	EM
				10%	KC	KC	KC	FC
				20%	KC	HC	FC	FC
				30%	KC	HC	FC	FC
				40%	KC,FC	HC	EM	FC
			100	1%	EM	EM	KC,FC	FC
				10%	EM	EM	FC	FC
				20%	KC	HC	FC	FC
				30%	FC	HC	FC	FC
				40%	FC	HC	EM	FC
			300	1%	EM	EM	KC	KC,FC
				10%	EM	EM	FC	FC
				20%	EM	HC	FC	FC
				30%	KC,FC	HC	FC	FC
				40%	FC	HC	EM	FC

จากตารางที่ 5.1.1

- กรณีที่ฐานข้อมูลอยู่ในรูปแบบวงรี (Non Spherical) เมื่อใช้วิธี Calinski and Harabasz index (Pseudo F) เป็นเกณฑ์ในการวัดประสิทธิภาพของวิธีการจัดกลุ่มข้อมูล พบว่า กรณีค่าเฉลี่ยของการซ้อนทับ $(\bar{\omega})$ มีค่าน้อย วิธีการจัดกลุ่มแบบอัลกอริทึม EM มีประสิทธิภาพการจัดกลุ่มข้อมูลที่ดี แต่เมื่อค่าเฉลี่ยของการซ้อนทับ $(\bar{\omega})$ มีค่าเพิ่มขึ้น วิธีการจัดกลุ่มข้อมูลที่ดีโดยส่วนใหญ่จะเป็นวิธีการจัดกลุ่มแบบเคมีน และวิธีการจัดกลุ่มแบบฟัซซี่ซีมีน และเมื่อใช้วิธี Silhouette width เป็นเกณฑ์ในการวัดประสิทธิภาพของวิธีการจัดกลุ่มข้อมูล พบว่า กรณีค่าเฉลี่ยของการซ้อนทับ $(\bar{\omega})$ มีค่าน้อย วิธีการจัดกลุ่มแบบอัลกอริทึม EM มีประสิทธิภาพการจัดกลุ่มข้อมูลที่ดี แต่เมื่อค่าเฉลี่ยของการซ้อนทับ $(\bar{\omega})$ มีค่าเพิ่มขึ้น วิธีการจัดกลุ่มแบบลำดับชั้น เป็นวิธีการจัดกลุ่มข้อมูลที่มีประสิทธิภาพที่ดี

- กรณีที่ฐานข้อมูลอยู่ในรูปแบบวงกลม (Spherical or Isotopic) เมื่อใช้วิธี Calinski and Harabasz index (Pseudo F) เป็นเกณฑ์ในการวัดประสิทธิภาพของวิธีการจัดกลุ่มข้อมูล พบว่า วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพที่ แต่ในกรณีที่ค่าเฉลี่ยของการซ้อนทับ $(\bar{\omega})$ มีค่าเท่ากับ 40% วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพที่ดี เมื่อใช้วิธี Silhouette width เป็นเกณฑ์ในการวัดประสิทธิภาพของวิธีการจัดกลุ่มข้อมูล พบว่า วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน มีประสิทธิภาพการจัดกลุ่มข้อมูลที่ดี

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 5.1.2 สรุปผลวิธีการจัดกลุ่มข้อมูลด้วยค่า Pseudo F และค่า Silhouette width โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 2 กลุ่ม จำนวนตัวแปรที่ใช้ศึกษา 3 ตัวแปร

K	p	k	n	$\bar{\omega}$	วิธีการจัดกลุ่ม			
					NonSph		Sph	
					CH	Sil	CH	Sil
2	3	2	50	1%	EM	FC,EM	KC	FC
				10%	KC,FC	FC	FC	FC
				20%	KC	HC	FC	FC
				30%	KC	FC	FC	FC
				40%	FC	FC	FC	FC
			100	1%	EM	EM	FC	EM
				10%	FC	FC	FC	FC
				20%	FC	HC	FC	FC
				30%	FC	HC	FC	FC
				40%	FC	HC	FC	FC
			300	1%	EM	EM	FC	EM
				10%	EM	EM	FC	FC
				20%	EM	HC	FC	FC
				30%	KC	HC	FC	FC
				40%	FC	HC	FC	FC

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จากตารางที่ 5.1.2

- กรณีที่ฐานข้อมูลอยู่ในรูปแบบวงรี (Non Spherical) เมื่อใช้วิธี Calinski and Harabasz index (Pseudo F) เป็นเกณฑ์ในการวัดประสิทธิภาพของวิธีการจัดกลุ่มข้อมูล พบว่า กรณีค่าเฉลี่ยของการซ้อนทับ $(\bar{\omega})$ มีค่าน้อย วิธีการจัดกลุ่มแบบอัลกอริทึม EM มีประสิทธิภาพการจัดกลุ่มข้อมูลที่ดี แต่เมื่อค่าเฉลี่ยของการซ้อนทับ $(\bar{\omega})$ มีค่าเพิ่มขึ้น วิธีการจัดกลุ่มแบบเคมีน และวิธีการจัดกลุ่มแบบฟัซซี่ซีมีน มีประสิทธิภาพการจัดกลุ่มข้อมูลที่ดี และเมื่อใช้วิธี Silhouette width เป็นเกณฑ์ในการวัดประสิทธิภาพของวิธีการจัดกลุ่มข้อมูล พบว่า กรณีค่าเฉลี่ยของการซ้อนทับ $(\bar{\omega})$ มีค่าน้อย วิธีการจัดกลุ่มแบบอัลกอริทึม EM มีประสิทธิภาพการจัดกลุ่มข้อมูลที่ดี แต่เมื่อค่าเฉลี่ยของการซ้อนทับ $(\bar{\omega})$ มีค่าเพิ่มขึ้น วิธีการจัดกลุ่มแบบลำดับชั้น เป็นวิธีการจัดกลุ่มข้อมูลที่มีประสิทธิภาพที่ดีเมื่อข้อมูลมีขนาดตัวอย่างเพิ่มขึ้น

- กรณีที่ฐานข้อมูลอยู่ในรูปแบบวงกลม (Spherical or Isotopic) เมื่อใช้วิธี Calinski and Harabasz index (Pseudo F) และวิธี Silhouette width เป็นเกณฑ์ในการวัดประสิทธิภาพของวิธีการจัดกลุ่มข้อมูล พบว่า วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพที่ดี

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

5.1.2 จำลองข้อมูลกรณีที่ฐานข้อมูลอยู่ในรูปแบบวงรี (Non Spherical) และกรณีที่ฐานข้อมูลอยู่ในรูปแบบวงกลม (Spherical or Isotopic) โดยที่ข้อมูลมีอัตราการซ้อนทับเฉลี่ย $(\bar{\omega})$ ที่ 1%, 10%, 20%, 30% และ 40% จำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม มีจำนวนตัวแปรที่ใช้ศึกษา คือ 2 และ 3 ตัวแปร โดยจำลองข้อมูลขนาดกลุ่มละ 50, 100 และ 300 ตัวอย่าง กำหนดจำนวนกลุ่มของการจัดกลุ่มข้อมูลเป็น 3 กลุ่มตามการจำลอง เพื่อทำการเปรียบเทียบประสิทธิภาพวิธีการจัดกลุ่ม 4 วิธี โดยใช้วิธี Calinski and Harabasz index (Pseudo F) และวิธี Silhouette width เป็นเกณฑ์ในการวัดประสิทธิภาพของวิธีการจัดกลุ่มข้อมูล สามารถสรุปผลได้ดังนี้

ตารางที่ 5.1.3 สรุปผลวิธีการจัดกลุ่มข้อมูลด้วยค่า Pseudo F และค่า Silhouette width โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม จำนวนตัวแปรที่ใช้ในการศึกษา 2 ตัวแปร

K	p	k	n	$\bar{\omega}$	วิธีการจัดกลุ่ม			
					NonSph		Sph	
					CH	Sil	CH	Sil
3	2	3	50	1%	EM	FC	EM	FC
				10%	KC	FC	EM	FC
				20%	KC	KC	EM	FC
				30%	KC	FC	EM	FC
				40%	FC	FC	FC	EM
			100	1%	EM	FC	HC	FC
				10%	EM	FC	EM	EM
				20%	HC	FC	EM	EM
				30%	HC	FC	EM	EM
				40%	FC	FC	FC	EM
			300	1%	EM	FC	HC	FC
				10%	EM	KC	EM	EM
				20%	KC	FC	EM	EM
				30%	FC	FC	EM	EM
				40%	FC	FC	EM	EM

จากตารางที่ 5.1.3

- กรณีที่ฐานข้อมูลอยู่ในรูปแบบวงรี (Non Spherical) เมื่อใช้วิธี Calinski and Harabasz index (Pseudo F) เป็นเกณฑ์ในการวัดประสิทธิภาพของวิธีการจัดกลุ่มข้อมูล พบว่า กรณีค่าเฉลี่ยของการซ้อนทับ $(\bar{\omega})$ มีค่าน้อย วิธีการจัดกลุ่มแบบอัลกอริทึม EM มีประสิทธิภาพการจัดกลุ่มข้อมูลที่ดี แต่เมื่อค่าเฉลี่ยของการซ้อนทับ $(\bar{\omega})$ มีค่าเพิ่มขึ้น วิธีการจัดกลุ่มแบบเคมีน และวิธีการจัดกลุ่มแบบฟัซซี่ซีมีน มีประสิทธิภาพการจัดกลุ่มข้อมูลที่ดี และเมื่อใช้วิธี Silhouette width เป็นเกณฑ์ในการวัดประสิทธิภาพของวิธีการจัดกลุ่มข้อมูล พบว่า วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพที่ดี

- กรณีที่ฐานข้อมูลอยู่ในรูปแบบวงกลม (Spherical or Isotopic) เมื่อใช้วิธี Calinski and Harabasz index (Pseudo F) เป็นเกณฑ์ในการวัดประสิทธิภาพของวิธีการจัดกลุ่มข้อมูล พบว่า วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพที่ดี และเมื่อใช้วิธี Silhouette width เป็นเกณฑ์ในการวัดประสิทธิภาพของวิธีการจัดกลุ่มข้อมูล พบว่า ข้อมูลมีขนาดเล็ก วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพที่ดี แต่เมื่อขนาดข้อมูลเพิ่มขึ้น วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพที่ดี

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 5.1.4 สรุปผลวิธีการจัดกลุ่มข้อมูลด้วยค่า Pseudo F และค่า Silhouette width โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 3 กลุ่ม จำนวนจัวแปรที่ใช้ในการศึกษา 3 ตัวแปร

K	p	k	n	$\bar{\omega}$	วิธีการจัดกลุ่ม			
					NonSph		Sph	
					CH	Sil	CH	Sil
3	3	3	50	1%	FC	FC	EM	EM
				10%	FC	FC	EM	EM
				20%	FC	KC	EM	EM
				30%	EM	FC	EM	EM
				40%	EM	FC	EM	EM
			100	1%	EM	FC	EM	EM
				10%	FC	FC	EM	EM
				20%	FC	FC	EM	EM
				30%	FC	EM	EM	EM
				40%	EM	EM	EM	EM
			300	1%	FC	FC	EM	EM
				10%	EM	EM	EM	EM
				20%	FC	FC	EM	EM
				30%	KC	FC	EM	EM
				40%	FC	KC	EM	EM

จากตารางที่ 5.1.4

- กรณีที่ฐานข้อมูลอยู่ในรูปแบบวงรี (Non Spherical) เมื่อใช้วิธี Calinski and Harabasz index (Pseudo F) และวิธี Silhouette width เป็นเกณฑ์ในการวัดประสิทธิภาพของวิธีการจัดกลุ่มข้อมูล พบว่า วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน และวิธีการจัดกลุ่มแบบอัลกอริทึม EM มีประสิทธิภาพการจัดกลุ่มข้อมูลที่ดี

- กรณีที่ฐานข้อมูลอยู่ในรูปแบบวงกลม (Spherical or Isotopic) เมื่อใช้วิธี Calinski and Harabasz index (Pseudo F) และวิธี Silhouette width เป็นเกณฑ์ในการวัดประสิทธิภาพของวิธีการจัดกลุ่มข้อมูล พบว่า วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพที่ดี

5.1.3 จำลองข้อมูลกรณีที่ฐานข้อมูลอยู่ในรูปแบบวงรี (Non Spherical) และกรณีที่ฐานข้อมูลอยู่ในรูปแบบวงกลม (Spherical or Isotopic) โดยที่ข้อมูลมีอัตราการซ้อนทับเฉลี่ย $(\bar{\omega})$ ที่ 1%, 10%, 20%, 30% และ 40% จำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม มีจำนวนตัวแปรที่ใช้ศึกษา คือ 2 และ 3 ตัวแปร โดยจำลองข้อมูลขนาดกลุ่มละ 50, 100 และ 300 ตัวอย่าง กำหนดจำนวนกลุ่มของการจัดกลุ่มข้อมูลเป็น 4 กลุ่มตามการจำลอง เพื่อทำการเปรียบเทียบประสิทธิภาพวิธีการจัดกลุ่ม 4 วิธี โดยใช้วิธี Calinski and Harabasz index (Pseudo F) และวิธี Silhouette width เป็นเกณฑ์ในการวัดประสิทธิภาพของวิธีการจัดกลุ่มข้อมูล สามารถสรุปผลได้ดังนี้

ตารางที่ 5.1.5 สรุปผลวิธีการจัดกลุ่มข้อมูลด้วยค่า Pseudo F และค่า Silhouette width โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม จำนวนตัวแปรศึกษา 2 ตัวแปร

K	p	k	n	$\bar{\omega}$	วิธีการจัดกลุ่ม			
					NonSph		Sph	
					CH	Sil	CH	Sil
4	2	4	50	1%	HC	KC	HC	EM
				10%	HC	EM	EM	EM
				20%	HC	EM	EM	EM
				30%	EM	EM	EM	EM
				40%	EM	EM	EM	EM
			100	1%	HC	FC	HC	EM
				10%	HC	EM	EM	EM
				20%	HC	EM	EM	EM
				30%	EM	EM	EM	EM
				40%	EM	EM	EM	EM
			300	1%	HC	KC	HC	EM
				10%	FC	EM	EM	EM
				20%	HC	EM	EM	EM
				30%	HC	EM	EM	EM
				40%	EM	EM	EM	EM

จากตารางที่ 5.1.5

- กรณีที่ฐานข้อมูลอยู่ในรูปแบบวงรี (Non Spherical) เมื่อใช้วิธี Calinski and Harabasz index (Pseudo F) เป็นเกณฑ์ในการวัดประสิทธิภาพของวิธีการจัดกลุ่มข้อมูล พบว่า กรณีค่าเฉลี่ยของการซ้อนทับ $(\bar{\omega})$ มีค่าน้อย วิธีการจัดกลุ่มแบบลำดับชั้น มีประสิทธิภาพการจัดกลุ่มข้อมูลที่ดี แต่เมื่อค่าเฉลี่ยของการซ้อนทับ $(\bar{\omega})$ มีค่าเพิ่มขึ้น วิธีการจัดกลุ่มแบบอัลกอริทึม EM มีประสิทธิภาพการจัดกลุ่มข้อมูลที่ดี และเมื่อใช้วิธี Silhouette width เป็นเกณฑ์ในการวัดประสิทธิภาพของวิธีการจัดกลุ่มข้อมูล พบว่า วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพที่ดี

- กรณีที่ฐานข้อมูลอยู่ในรูปแบบวงกลม (Spherical or Isotopic) เมื่อใช้วิธี Calinski and Harabasz index (Pseudo F) และวิธี Silhouette width เป็นเกณฑ์ในการวัดประสิทธิภาพของวิธีการจัดกลุ่มข้อมูล พบว่า วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพที่ดี

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 5.1.6 สรุปผลวิธีการจัดกลุ่มข้อมูลด้วยค่า Pseudo F และค่า Silhouette width โดยข้อมูลมีจำนวนกลุ่มของการซ้อนทับกัน 4 กลุ่ม

K	p	k	n	$\bar{\omega}$	วิธีการจัดกลุ่ม			
					NonSph		Sph	
					CH	Sil	CH	Sil
4	3	4	50	1%	EM	KC	EM	EM
				10%	EM	EM	EM	EM
				20%	EM	EM	EM	EM
				30%	EM	EM	EM	EM
				40%	EM	KC	EM	EM
			100	1%	EM	FC	EM	EM
				10%	HC	KC,FC	EM	EM
				20%	EM	EM	EM	EM
				30%	EM	EM	EM	EM
				40%	EM	EM	EM	EM
			300	1%	EM	FC	EM	EM
				10%	HC	KC	EM	EM
				20%	HC	KC	EM	EM
				30%	EM	KC	EM	EM
				40%	EM	EM	EM	EM

จากตารางที่ 5.1.6

- กรณีที่ฐานข้อมูลอยู่ในรูปแบบวงรี (Non Spherical) เมื่อใช้วิธี Calinski and Harabasz index (Pseudo F) เป็นเกณฑ์ในการวัดประสิทธิภาพของวิธีการจัดกลุ่มข้อมูล พบว่า วิธีการจัดกลุ่มแบบอัลกอริทึม EM มีประสิทธิภาพการจัดกลุ่มข้อมูลที่ดี และเมื่อใช้วิธี Silhouette width เป็นเกณฑ์ในการวัดประสิทธิภาพของวิธีการจัดกลุ่มข้อมูล พบว่า วิธีการจัดกลุ่มแบบฟัซซี่ซีมีน และวิธีการจัดกลุ่มแบบอัลกอริทึม EM มีประสิทธิภาพการจัดกลุ่มข้อมูลที่ดี

- กรณีที่ฐานข้อมูลอยู่ในรูปแบบวงกลม (Spherical or Isotopic) เมื่อใช้วิธี Calinski and Harabasz index (Pseudo F) และวิธี Silhouette width เป็นเกณฑ์ในการวัดประสิทธิภาพของวิธีการจัดกลุ่มข้อมูล พบว่า วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพที่ดี

<u>สรุป</u>

- กรณีที่ฐานข้อมูลอยู่ในรูปแบบวงรี (Non Spherical) เมื่อใช้วิธี Calinski and Harabasz index (Pseudo F) และวิธี Silhouette width เป็นเกณฑ์ในการวัดประสิทธิภาพของวิธีการจัดกลุ่มข้อมูล พบว่า วิธีการจัดกลุ่มข้อมูลทั้ง 4 วิธี เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพที่ดี ตามแต่ละสถานการณ์

- กรณีที่ฐานข้อมูลอยู่ในรูปแบบวงกลม (Spherical or Isotopic) เมื่อใช้วิธี Calinski and Harabasz index (Pseudo F) และวิธี Silhouette width เป็นเกณฑ์ในการวัดประสิทธิภาพของวิธีการจัดกลุ่มข้อมูล พบว่า วิธีการจัดกลุ่มแบบอัลกอริทึม EM เป็นวิธีการจัดกลุ่มที่มีประสิทธิภาพที่ดี เมื่อจำนวนกลุ่มของการซ้อนทับ และอัตราการซ้อนทับเฉลี่ย $(\bar{\omega})$ มีค่าเพิ่มขึ้น

5.2 ข้อเสนอแนะ

สำหรับผู้ที่สนใจศึกษาเพิ่มเติม ในการศึกษาครั้งต่อไป อาจทำการศึกษาในกรณีต่าง ๆ ดังต่อไปนี้

1. ในงานวิจัยชิ้นนี้ ผู้วิจัยได้ศึกษาเฉพาะกรณีที่ข้อมูลมีการแจกแจงปกติ และมีตัวแปรศึกษาเพียง 2 และ 3 ตัวแปร สำหรับผู้ที่สนใจศึกษางานวิจัยครั้งนี้ต่ออาจทำการศึกษากรณีที่ข้อมูลมีรูปแบบการแจกแจงอื่น ๆ และตัวแปรศึกษาที่มากกว่านี้ อาจจะทำให้ผลการวิจัยที่ได้แตกต่างจากเดิม

2. ศึกษาดัชนีการตัดสินใจในแต่ละวิธีการจัดกลุ่มว่าข้อมูลเพิ่มเติม นอกเหนือจาก 2 วิธี ที่ผู้วิจัยได้ทำการศึกษามาก่อนแล้ว

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

รายการอ้างอิง

Bezdek, J. C. (1981). "Pattern Recognition with Fuzzy Objective Function Algoritms." Plenum Press, New York.

Calinski, T. and J. Harabasz (1974). "A Dendrite Method for Cluster Analysis." Communications in Statistics 3: 1-27.

Dempster, A. P., et al. (1977). "Maximum Likelihood from Incomplete Data via the EM Algorithm." Journal of the Royal Statistical Society(1): 1-38.

Dunn, J. C. (1973). "A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters." Journal of Cybernetics 3: 32-57.

Johnson, S. C. (1967). "Hierarchical Clustering Schemes." Psychometrika 2: 241-254.

MacQueen, J. B. (1967). "Some Methods for classification and Analysis of Multivariate Observations, Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability." Berkeley,University of California Press(1): 281-297.

Melnykov, V., et al. (2012). "An R Package for Simulating Data to Study Performance of Clustering Algorithms." Journal of Statistical Software 51(12).

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ภาคผนวก

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

รายละเอียดของโปรแกรมที่ใช้ในงานวิจัย

การศึกษาวิจัยครั้งนี้ใช้โปรแกรม R เวอร์ชั่น 3.0.2 ในการจำลองข้อมูลและการจัดกลุ่มข้อมูล และการคำนวณค่าเพื่อทำการเปรียบเทียบวิธีการจัดกลุ่ม ซึ่งมีคำสั่งดังต่อไปนี้

```
###### Non Spherical ######
        Data <- MixSim(BarOmega, MaxOmega = NULL, K, p, sph = FALSE, hom = FALSE)
        nData <- simdataset(n, Pi = Data$Pi, Mu = Data$Mu, S = Data$S)

###### Non Spherical ######
        Data <- MixSim(BarOmega, MaxOmega = NULL, K, p, sph = TRUE, hom = FALSE)
        nData <- simdataset(n, Pi = Data$Pi, Mu = Data$Mu, S = Data$S)

### Hierarchical Cluster ###
        HC <- hclust(dist(Dataset), "ward")
        CutHC <- cutree(HC,k)
        HCCluststat <- cluster.stats(dist(Dataset),CutHC)
        HCPFI <- HCCluststat$ch
        HCsil <- silhouette(CutHC,dist(Dataset))
        HCSIL <- summary(HCsil)$avg.width

### K-means Cluster ###
        KC <- kmeans(Dataset,k)
        KCCluststat <- cluster.stats(dist(Dataset),KC$cluster)
        KCPFI <- KCCluststat$ch
        KCsil <- silhouette(KC$cluster,dist(Dataset))
        KCSIL <- summary(KCsil)$avg.width

### Fuzzy c-means Cluster ###
        FC <- cmeans(Dataset,k)
        FCCluststat <- cluster.stats(dist(Dataset),FC$cluster)
        FCPFI <- FCCluststat$ch
        FCsil <- silhouette(FC$cluster,dist(Dataset))
        FCSIL <- summary(FCsil)$avg.width

 ### EM Cluster ###
        EM <- Mclust(Dataset,k)
        EMCluststat <- cluster.stats(dist(Dataset),EM$classification)
        EMPFI <- EMCluststat$ch
        EMsil <- silhouette(EM$classification,dist(Dataset))
        EMSIL <- summary(EMsil)$avg.width
```

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ประวัติผู้เขียนวิทยานิพนธ์

นางสาวจิรวรรณ ไพบูลย์วรชาติ เกิดวันศุกร์ที่ 24 พฤศจิกายน พ.ศ.2532 สำเร็จการศึกษาปริญญาวิทยาศาสตรบัณฑิต (วท.บ.) สาขาสถิติ ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ มหาวิทยาลัยบูรพา ในปีการศึกษา 2554 และเข้าศึกษาต่อในหลักสูตรวิทยาศาสตรมหาบัณฑิต (วท.ม.) สาขาวิชาสถิติ ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2555

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University