มาตรวัดระยะห่างสำหรับข้อมูลแบบผสมกับการวิเคราะห์กลุ่ม

นางสาวพิชญา บุตรขุนทอง

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาสถิติ ภาควิชาสถิติ
คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2558

DISTANCE MEASURES FOR MIXED DATA WITH APPLICATION IN CLUSTER ANALYSIS

Miss Pitchaya Buthkhunthong

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 2015

หัวข้อวิทยานิพนธ์	มาตรวัดระยะห่างสำหรับข้อมูลแบบผสมกับการวิเคราะห์กลุ่ม
โดย	นางสาวพิชญา บุตรขุนทอง
สาขาวิชา	สถิติ
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	อาจารย์ ดร. อัครินทร์ ไพบูลย์พานิช

คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

.....คณบดีคณะพาณิชยศาสตร์และการบัญชี

(รองศาสตราจารย์ ดร. พสุ เดชะรินทร์)

คณะกรรมการสอบวิทยานิพนธ์

.....ประธานกรรมการ

(รองศาสตราจารย์ ดร. เสกสรร เกียรติสุไพบูลย์)

.....อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

(อาจารย์ ดร. อัครินทร์ ไพบูลย์พานิช)

.....กรรมการ

(อาจารย์ ดร. นัท กุลวานิช)

.....กรรมการภายนอกมหาวิทยาลัย

(อาจารย์ ดร. อรุณี กำลัง)

พิชญา บุตรขุนทอง : มาตรวัดระยะห่างสำหรับข้อมูลแบบผสมกับการวิเคราะห์กลุ่ม (DISTANCE MEASURES FOR MIXED DATA WITH APPLICATION IN CLUSTER ANALYSIS) อ.ที่ปรึกษาวิทยานิพนธ์หลัก: อ. ดร. อัครินทร์ ไพบูลย์พานิช, 185 หน้า.

การศึกษาวิจัยนี้ได้เปรียบเทียบประสิทธิภาพการวิเคราะห์กลุ่มข้อมูลแบบผสม ซึ่งประกอบไปด้วยตัวแปรนามบัญญัติ ตัวแปรอันดับ และตัวแปรเชิงปริมาณ ด้วยอัลกอริทึมจัดกลุ่มโดยรอบมีดอยด์ โดยใช้มาตรวัดระยะห่างแบบต่าง ๆ คือ ระยะห่างของ Kaufman and Rousseeuw (KR) ระยะห่างของ Podani (P) ซึ่งทั้งสองพัฒนามาจากความคล้ายของ Gower และมาตรวัดระยะห่างที่เสนอขึ้นใหม่โดยประยุกต์ระยะห่างสำหรับตัวแปรนามบัญญัติของ Noorbehbahani et al. (N) ร่วมกับระยะห่างของ KR และระยะห่างของ P นั่นคือระยะห่างแบบ KR&N และระยะห่างแบบ P&N โดยจำลองข้อมูลแบบผสมและข้อมูลรูปแบบอื่น ๆ ที่ประกอบไปด้วยตัวแปรต่างชนิดกัน และกำหนดให้ทราบกลุ่มแน่ชัด โดยศึกษาภายใต้ขอบเขตค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรเท่ากับ 0.2 และ 0.8 ขนาดข้อมูลต่อกลุ่มเท่ากับ 20 และ 100 จำนวนกลุ่มข้อมูลเท่ากับ 3 และ 5 จำนวนประเภทของตัวแปรนามบัญญัติและจำนวนอันดับของตัวแปรอันดับเท่ากับ 5 และพิจารณากรณีที่ความถี่ของข้อมูลแต่ละประเภทหรืออันดับแตกต่างและไม่แตกต่างกัน

ผลการศึกษาพบว่า กรณีที่ความถี่ของข้อมูลแต่ละประเภทหรืออันดับแตกต่างกัน สำหรับข้อมูลแบบผสมที่ประกอบไปด้วยตัวแปรทั้ง 3 ชนิด การวิเคราะห์กลุ่มด้วยระยะห่างแบบ KR&N มีประสิทธิภาพดีที่สุด และการวิเคราะห์กลุ่มด้วยระยะห่างแบบ P&N มีประสิทธิภาพรองลงมา นอกจากนี้ระยะห่างแบบ KR&N เหมาะสำหรับการวิเคราะห์กลุ่มข้อมูลที่ประกอบไปด้วยทั้งตัวแปรนามบัญญัติและตัวแปรเชิงปริมาณ ขณะที่ระยะห่างของ KR เหมาะสำหรับการวิเคราะห์กลุ่มข้อมูลที่ประกอบไปด้วยทั้งตัวแปรอันดับและตัวแปรเชิงปริมาณ อย่างไรก็ตามกรณีที่ความถี่ของข้อมูลแต่ละประเภทหรืออันดับไม่แตกต่างกัน พบว่า โดยส่วนใหญ่การวิเคราะห์กลุ่มข้อมูลด้วยระยะห่างแบบต่าง ๆ มีประสิทธิภาพไม่แตกต่างกัน

ภาควิชา สถิติ ลายมือชื่อนิสิต .....

สาขาวิชา สถิติ ลายมือชื่อ อ.ที่ปรึกษาหลัก .....

ปีการศึกษา 2558

5681567826 : MAJOR STATISTICS
KEYWORDS: NOMINAL VARIABLE / ORDINAL VARIABLE / QUANTITATIVE VARIABLE / MIXED DATA / DISTANCE / CLUSTER

PITCHAYA BUTHKHUNTHONG: DISTANCE MEASURES FOR MIXED DATA WITH APPLICATION IN CLUSTER ANALYSIS. ADVISOR: AKARIN PHAIBULPANICH, 185 pp.

This study presents performance comparison of cluster analysis through Partitioning Around Medoids algorithm, for mixed data which contains nominal, ordinal, and numerical variables, using different types of distance measures: Kaufman and Rousseeuw distance (KR) and Podani distance (P) which are adapted from Gower's similarity, and two newly proposed distance measures: one is a combination between KR and Noorbehbahani et al. distance (KR&N) and the other is a combination between P and Noorbehbahani et al. distance (P&N). Mixed data and other types of data were simulated with equal and unequal frequency of nominal and ordinal variables. This study also sets correlations between variables at 0.2 and 0.8, 20 and 100 instances per group, 3 and 5 groups, and 5 values of nominal and ordinal variables.

In case of unequal frequency data, the clustering using KR&N distance gives better result for mixed data. Moreover, the clustering using KR&N distance is suitable for the data which contains both nominal and numerical variables, while the clustering using KR distance is suitable for the data which contains only ordinal and numerical variables. However, in case of equal frequency data, four distances show similar efficiency.

Department:	Statistics	Student's Signature	.....
Field of Study:	Statistics	Advisor's Signature	.....
Academic Year:	2015		

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้เสร็จสมบูรณ์ลงได้ด้วยความช่วยเหลือ และการเอาใจใส่จากอาจารย์ที่ปรึกษาวิทยานิพนธ์ อาจารย์ ดร. อัครินทร์ ไพบูลย์พานิช ผู้วิจัยจึงขอกราบขอบพระคุณท่านอาจารย์เป็นอย่างสูงที่กรุณาให้คำปรึกษา ชี้แนะแนวทาง อบรมสั่งสอน ให้ความช่วยเหลือ และเป็นกำลังใจที่ดีเสมอมา

ผู้วิจัยขอกราบขอบพระคุณท่านประธานในการสอบวิทยานิพนธ์ รองศาสตราจารย์ ดร. เสกสรร เกียรติสุไพบูลย์ ท่านกรรมการ อาจารย์ ดร. นัท กุลวานิช และท่านกรรมการภายนอกมหาวิทยาลัย อาจารย์ ดร. อรุณี กำลัง เป็นอย่างสูง ที่ท่านอาจารย์ทั้งสามท่านได้เสียสละเวลาเพื่อสอบ ตรวจสอบ และให้คำแนะนำที่ดี เพื่อปรับปรุงวิทยานิพนธ์ฉบับนี้ให้สมบูรณ์ยิ่งขึ้น อีกทั้งขอกราบขอบพระคุณอาจารย์ทุกท่านประจำภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย ที่ได้ประสิทธิ์ประสาทวิชาความรู้ให้ผู้วิจัย ทำให้ผู้วิจัยสามารถนำความรู้เหล่านั้นมาใช้ในวิทยานิพนธ์ได้อย่างเต็มที่

สุดท้ายนี้ผู้วิจัยขอขอบพระคุณ คุณพ่อ คุณแม่ ครอบครัว รวมทั้งเพื่อน ๆ ที่คอยช่วยเหลือ สนับสนุน ส่งเสริม และเป็นกำลังใจในเรื่องต่าง ๆ ให้แก่ผู้วิจัยตลอดมา

สารบัญ

หน้า

หน้า

หน้า

สารบัญตาราง

หน้า

สารบัญรูป

หน้า

บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

การวิเคราะห์กลุ่ม (Cluster analysis) เป็นการวิเคราะห์แบ่งกลุ่มข้อมูลหรือจำแนกข้อมูล ซึ่งข้อมูลที่มีลักษณะคล้ายคลึงกันจะถูกจัดให้อยู่ในกลุ่มเดียวกัน และข้อมูลที่มีลักษณะแตกต่างกันจะถูกจัดให้อยู่ต่างกลุ่มกัน การวิเคราะห์กลุ่มเป็นเทคนิคที่ถูกนำไปใช้ในหลายแขนงวิชา ทั้งในสังคมศาสตร์ ภูมิศาสตร์ เศรษฐศาสตร์ หลักการ ตลาด และชีววิทยา เป็นต้น ในทางปฏิบัติ ข้อมูลจริงอาจเป็นทั้งตัวแปรเชิงปริมาณ (Quantitative variable) หรือ ตัวแปรเชิงคุณภาพ (Qualitative variable) ซึ่งแบ่งได้อีกเป็นตัวแปรนามบัญญัติ (Nominal variable) และตัวแปรอันดับ (Ordinal variable) เช่น การจำแนกกลุ่มผู้ตอบแบบสอบถามเกี่ยวกับการซื้อรถยนต์ โดยพิจารณาจาก รายได้ อายุ อาชีพ และระดับการศึกษา เพื่อหาแนวทางการพัฒนารถยนต์ให้ตรงตามความต้องการเฉพาะกลุ่ม

การพิจารณาว่าข้อมูลสองข้อมูลคล้ายคลึงกันหรือแตกต่างกัน ต้องอาศัยมาตรวัดความคล้าย (Similarity measure) หรือมาตรวัดระยะห่าง (Distance measure) ระหว่างจุดข้อมูลแต่ละคู่ (Everitt, Landau et al. 2011) ทั้งนี้มาตรวัดโดยส่วนใหญ่ เช่น ระยะห่างแบบยูคลิเดียน (Euclidean distance) และระยะห่างแบบแมนฮัตตัน (Manhattan distance) จะใช้ได้เฉพาะข้อมูลที่ประกอบไปด้วยตัวแปรเชิงปริมาณ ขณะที่มาตรวัดสำหรับข้อมูลเชิงคุณภาพนั้น มีมาตรวัดที่ถูกคิดค้นขึ้นมาพอสมควร อาทิ ดัชนีแจคการ์ด (Jaccard index) ซึ่งจะพิจารณาว่าระหว่างจุดข้อมูลคู่นั้น ๆ มีตัวแปรใดบ้างที่มีค่าต่างประเภทกัน แต่ใช้ได้เฉพาะตัวแปรทวินาม (Binary variable) หรือตัวแปรดัมมี่ (Dummy variable) ซึ่งเป็นตัวแปรนามบัญญัติที่มีเพียงค่า 0 และ 1 เท่านั้น

อย่างไรก็ดี ในทางปฏิบัติ ข้อมูลอาจประกอบไปด้วยทั้งตัวแปรเชิงปริมาณ ตัวแปรนามบัญญัติ และตัวแปรอันดับ นั่นคือข้อมูลแบบผสม (Mixed data) ดังเช่นตัวอย่างการแบ่งกลุ่มผู้ตอบแบบสอบถามเกี่ยวกับการซื้อรถยนต์ที่ได้กล่าวก่อนหน้านี้ ซึ่งประกอบไปด้วยข้อมูลรายได้ อายุ ซึ่งเป็นตัวแปรเชิงปริมาณ ข้อมูลด้านอาชีพ ซึ่งเป็นตัวแปรนามบัญญัติ และข้อมูลระดับการศึกษา ซึ่งเป็นตัวแปรอันดับ จะเห็นได้ว่าหากตัดตัวแปรใดไป กลุ่มของลูกค้าที่ได้จากการวิเคราะห์กลุ่มจะขาดลักษณะของตัวแปรนั้น ๆ ส่งผลให้การกำหนดแนวทางการพัฒนารถยนต์ให้สอดคล้องกับลักษณะที่แท้จริงของกลุ่มและตรงตามความต้องการเฉพาะกลุ่มของลูกค้าเป็นไปได้ยาก นั่นคือการคัดเลือกตัวแปรในแต่ละชุดข้อมูลสามารถส่งผลต่อประสิทธิภาพที่ดีของการวิเคราะห์กลุ่ม ทำให้ในหลายสถานการณ์ไม่สามารถหลีกเลี่ยงการใช้ตัวแปรเชิงคุณภาพทั้งสองชนิดในการวิเคราะห์กลุ่มร่วมกับตัวแปรเชิงปริมาณได้

ในการวัดความต่างสำหรับตัวแปรนามบัญญัติ วิธีการหนึ่งก็คือการแปลงให้เป็นตัวแปรดัมมี่แล้วจึงใช้มาตรวัดแบบต่าง ๆ เช่นเดียวกับตัวแปรดัมมี่ หากตัวแปรนั้น ๆ มีจำนวนประเภทค่อนข้างมากก็จะทำให้ได้ตัวแปรดัมมี่มีจำนวนมากตามไปด้วย เช่น อาชีพของลูกค้า อาจสามารถจำแนกเป็นประเภทใหญ่ ๆ ได้ 10 ประเภท คือ ข้าราชการ พนักงานรัฐวิสาหกิจ พนักงานบริษัท ธุรกิจส่วนตัว ค้าขาย รับจ้าง/ลูกจ้าง นิสิต/นักศึกษา เกษตรกร/ปศุสัตว์/ประมง เกษียณ/ว่างงาน และอื่น ๆ ซึ่งจะต้องใช้ตัวแปรดัมมี่มากถึง 9 ตัวแปร เป็นต้น ทั้งนี้จำนวนตัวแปรที่มากเกินไป ทำให้จำนวนมิติของข้อมูลเพิ่มมากขึ้น ทั้งยังทำให้ใช้เวลามากขึ้นในการวิเคราะห์กลุ่ม นอกจากนี้เมื่อพิจารณาข้อมูลที่มีทั้งตัวแปรเชิงปริมาณและตัวแปรนามบัญญัติ ระยะห่างของข้อมูลจะต้องวัดจากตัวแปรทั้งสองชนิด โดยตัวแปรเชิงปริมาณไม่มีค่าขอบเขตที่แน่นอน ขณะที่ตัวแปรนามบัญญัติถูกแปลงเป็นตัวแปรดัมมี่จะมีค่าเพียง 0 หรือ 1 เท่านั้น หากใช้การวัดระยะห่างแบบยูคลิเดียนหรือแมนฮัตตันโดยตรง น้ำหนักของแต่ละตัวแปรในการคำนวณระยะห่างจะไม่สมดุล เช่น ลูกค้าที่จะซื้อรถยนต์ 2 คน มีผลต่างอายุเท่ากับ 15 ปี แต่ประกอบอาชีพต่างกัน ผลต่างของตัวแปรอาชีพจึงเท่ากับ 1 ดังนั้นลูกค้าแต่ละคนจะต่างจากลูกค้าคนอื่น เนื่องจากตัวแปรอายุเป็นหลัก ทำให้ไม่ได้ระยะห่างที่ใกล้เคียงความเป็นจริงนั่นเอง

ในการวัดความต่างสำหรับตัวแปรอันดับ วิธีการหนึ่งก็คืออาจทำการแปลงตัวแปรอันดับเป็นตัวแปรดัมมี่ เช่นเดียวกับตัวแปรนามบัญญัติ แต่วิธีการนี้นอกจากจะทำให้เกิดตัวแปรจำนวนมากเช่นเดียวกับกรณีตัวแปรนามบัญญัติแล้ว ยังทำให้เกิดการสูญเสียอันดับของข้อมูลอีกด้วย ทั้งยังไม่เหมาะสมที่จะใช้มาตรวัดความต่างแบบเดียวกับตัวแปรเชิงปริมาณ เนื่องจากตัวแปรอันดับบอกเพียงลำดับของข้อมูลว่ามากหรือน้อยกว่ากัน แต่ไม่สามารถระบุระยะห่างที่แท้จริงของแต่ละระดับได้ เช่น ไม่ทราบว่าการศึกษาระดับปริญญาตรีต่างจากระดับปริญญาโทมากเท่าใด แต่สามารถทราบได้ว่าอายุของลูกค้าแต่ละคนต่างกันกี่ปี ดังนั้นมาตรวัดความต่างของตัวแปรแบบผสมที่ประกอบไปด้วยตัวแปรเชิงปริมาณ ตัวแปรนามบัญญัติ และตัวแปรอันดับ จึงถูกนิยามขึ้นใหม่อย่างหลากหลาย เพื่อให้สามารถใช้วิเคราะห์กลุ่มได้อย่างมีประสิทธิภาพมากที่สุด

Gower (1971) ได้เสนอมาตรวัดความคล้ายสำหรับข้อมูลที่ประกอบไปด้วยตัวแปรเชิงปริมาณและตัวแปรนามบัญญัติ โดยใช้หลักค่าสัมประสิทธิ์การจับคู่อย่างง่าย (Simple matching coefficient) กับตัวแปรนามบัญญัติ นั่นคือข้อมูลคู่ใดที่มีค่าตัวแปรนามบัญญัติอยู่ในประเภทเดียวกันจะมีค่าความคล้ายเท่ากับ 1 หากต่างประเภทกัน จะมีค่าความคล้ายเท่ากับ 0 และนิยามความคล้ายสำหรับตัวแปรเชิงปริมาณขึ้นใหม่ โดยพิจารณาจากผลต่างค่าตัวแปรของข้อมูลคู่นั้นและหารด้วยพิสัยของตัวแปรนั้น จากนั้นลบทั้งหมดด้วย 1 จะได้ว่าความคล้ายของแต่ละตัวแปรมีค่าอยู่ระหว่าง 0 ถึง 1 จึงสามารถหาความคล้ายของข้อมูลที่มีทั้งตัวแปรเชิงปริมาณและตัวแปรนามบัญญัติได้ อย่างไรก็ตาม Gower ไม่ได้กล่าวถึงความคล้ายสำหรับตัวแปรอันดับ ซึ่งอาจเป็นตัวแปรสำคัญในการวิเคราะห์กลุ่ม

Kaufman and Rousseeuw (1990) ได้ปรับเปลี่ยนมาตรวัดความคล้ายของ Gower ให้อยู่ในรูปของระยะห่าง รวมถึงเสนอมาตรวัดความต่างสำหรับตัวแปรอันดับเพิ่มเติม เพื่อให้สามารถวัดระยะห่างข้อมูลแบบผสมได้ โดยแปลงข้อมูลที่เป็นตัวแปรอันดับให้มีค่าอยู่ระหว่าง 0 ถึง 1 และนำค่าที่ถูกแปลงมาคำนวณหาระยะห่างเช่นเดียวกับตัวแปรเชิงปริมาณ ซึ่งในปัจจุบันถูกใช้อย่างแพร่หลายในรูปแบบของฟังก์ชัน daisy ในแพ็กเกจ cluster และ ฟังก์ชัน gower.dist ในแพ็กเกจ StatMatch ในโปรแกรม R เนื่องจากสามารถคำนวณระยะห่างสำหรับข้อมูลแบบผสมได้ อย่างไรก็ตามระยะห่างสำหรับตัวแปรอันดับของ Kaufman and Rousseeuw ที่ถูกแปลงยังคงมีระยะห่างเท่า ๆ กัน อาทิ อันดับ 1 2 3 4 และ 5 จะถูกแปลงเป็น 0 0.25 0.5 0.75 และ 1 ตามลำดับ ซึ่งระยะห่างแต่ละระดับจะมีค่าเท่ากับ 0.25 เท่า ๆ กัน เมื่อคำนวณด้วยสมการเดียวกับตัวแปรเชิงปริมาณจะไม่ต่างจากการนำอันดับของข้อมูลนั้นมาคำนวณด้วยสมการเดียวกับตัวแปรเชิงปริมาณโดยตรง ซึ่งไม่เหมาะสม เนื่องจากตัวแปรอันดับต่างจากตัวแปรเชิงปริมาณ คือระยะห่างของแต่ละระดับอาจมีปริมาณไม่เท่ากันก็ได้ดังที่กล่าวไว้ข้างต้น

Podani (1999) ได้สร้างมาตรวัดความคล้ายสำหรับตัวแปรอันดับเพิ่มเติมจากวิธีเดิมของ Gower โดยแปลงข้อมูลตัวแปรอันดับจากการเรียงข้อมูลทั้งหมดจากน้อยไปมากของตัวแปรนั้นให้มีอันดับตั้งแต่ 1 ถึงจำนวนข้อมูลทั้งหมด จากนั้นรวมค่าอันดับใหม่ที่เดิมมีอันดับเดียวกันทั้งหมด แล้วจึงหารด้วยจำนวนการซ้ำกันของอันดับนั้น ๆ กำหนดให้ใช้ค่าอันดับใหม่หรือเรียกว่าคะแนนอันดับ (Rank score) แทนค่าอันดับเริ่มต้นและนำไปคำนวณด้วยความคล้ายซึ่ง Podani ได้สร้างขึ้นใหม่ 2 วิธี สำหรับวิธีแรก คือการใช้สมการเดียวกับตัวแปรเชิงปริมาณของ Gower เมื่อแปลงเป็นค่าความต่างแล้ว จะได้เมทริกซ์ระยะห่าง (Distance matrix) ของข้อมูลที่มีคุณสมบัติเป็นเมทริก (Metric) ขณะที่วิธีที่สอง จะได้เมทริกซ์ระยะห่างที่ไม่มีคุณสมบัติเป็นเมทริก (Non-metric) วิธีนี้จึงไม่สามารถนำไปใช้กับเทคนิคการวิเคราะห์กลุ่มที่ต้องใช้เมทริกซ์ระยะห่างที่มีคุณสมบัติเป็นเมทริกได้ ทั้งนี้ Podani พิจารณามาตรวัดความคล้ายที่นำเสนอกับข้อมูลจริงเพียงชุดเดียวเท่านั้น และไม่มีการวัดประสิทธิภาพในการวิเคราะห์กลุ่มด้วยค่าตัวเลขที่ชัดเจน จึงเป็นที่สงสัยว่ามาตรวัดความคล้ายนี้ส่งผลให้การวิเคราะห์กลุ่มสำหรับข้อมูลแบบผสมมีประสิทธิภาพเพียงใด

Noorbehbahani, Mousavi et al. (2015) เล็งเห็นว่ามาตรวัดความต่างของตัวแปรนามบัญญัติไม่ควรมีค่าเพียง 0 และ 1 เท่านั้น แต่ควรมีค่าที่ขึ้นอยู่กับความถี่ของค่าของตัวแปรนั้น ๆ ด้วย กล่าวคือสำหรับค่าที่ต่างกันของตัวแปรนามบัญญัติ ค่าของตัวแปรที่มีความถี่สูง (เกิดขึ้นบ่อย) สองค่าควรมีระยะห่างน้อย ขณะที่ค่าของตัวแปรที่มีความถี่ต่ำสองค่า หรือค่าของตัวแปรสองค่าที่ค่าหนึ่งมีความถี่ต่ำกับอีกค่ามีความถี่สูงควรมีระยะห่างที่สูง เนื่องจากโอกาสที่จะถูกจัดให้อยู่ในกลุ่มเดียวกันไม่มากนัก นอกจากนี้ Noorbehbahani et al. ยังเสนอเทคนิคการวิเคราะห์กลุ่มขึ้นใหม่โดยใช้ระยะห่างที่เสนอ และเปรียบเทียบประสิทธิภาพกับเทคนิคการวิเคราะห์กลุ่มอื่น ๆ ด้วยการวิเคราะห์ข้อมูลจริง

ที่ทราบกลุ่มแน่ชัด และคำนวณหาค่าความแม่นยำ ซึ่งเป็นค่าที่บอกว่าผลการวิเคราะห์กลุ่มที่ได้ถูกต้องตามความเป็นจริงเพียงใด อย่างไรก็ตาม Noorbehbahani et al. ไม่ได้มีการจำลองข้อมูลที่อาจเกิดขึ้นได้ใน สถานการณ์ต่าง ๆ และพิจารณาเพียงกรณีที่ข้อมูลประกอบไปด้วยตัวแปรเชิงปริมาณและตัวแปรนามบัญญัติเท่านั้น ซึ่งยังไม่ได้พิจารณากรณีที่มีตัวแปรอันดับร่วมด้วย นั่นคือยังไม่ได้พิจารณามาตรวัดความต่างของตัวแปรแบบผสมทั้งสามชนิด

จากตารางที่ 1.1 จะเห็นว่าระยะห่างของ Kaufman and Rousseeuw (KR) และระยะห่างของ Podani (P) สามารถวัดระยะห่างของข้อมูลแบบผสมที่ประกอบไปด้วยตัวแปรเชิงปริมาณ ตัวแปรนามบัญญัติ และตัวแปรอันดับได้ทั้งหมด แต่ Kaufman and Rousseeuw และ Podani ไม่ได้กล่าวถึงระยะห่างของตัวแปรนามบัญญัติที่มีความถี่ของค่าของตัวแปรนั้น ๆ มาเกี่ยวข้องดังเช่นระยะห่างของ Noorbehbahani et al.

ตารางที่ 1.1 มาตรวัดความคล้าย/ความต่างแบบต่าง ๆ กับลักษณะของข้อมูลที่แตกต่างกัน

มาตรวัดความคล้าย /ความต่าง ลักษณะข้อมูล	ความคล้ายของ Gower	ระยะห่างของ Kaufman and Rousseeuw	ระยะห่างของ Podani	ระยะห่างของ Noorbehbahani et al.
ตัวแปรเชิงปริมาณ	✓	✓	✓	✓
ตัวแปรนามบัญญัติ	✓	✓	✓	✓
ตัวแปรอันดับ		✓	✓	

นอกจากการเลือกใช้มาตรวัดความต่างสำหรับข้อมูลแบบผสมแล้ว ยังต้องพิจารณาถึงเทคนิคการวิเคราะห์กลุ่มที่เหมาะสำหรับข้อมูลแบบผสม และสามารถปรับใช้กับมาตรวัดความต่างแบบต่าง ๆ ได้ ซึ่งอัลกอริทึมหนึ่งที่เหมาะสมคืออัลกอริทึมจัดกลุ่มโดยรอบมีดอยด์ (Partitioning around medoids algorithm, PAM algorithm) เนื่องจากอัลกอริทึมจัดกลุ่มโดยรอบมีดอยด์ เป็นอัลกอริทึมจัดกลุ่มรูปแบบหนึ่งของอัลกอริทึมจัดกลุ่มแบบเคมีดอยด์ (K-Medoids clustering algorithm) ซึ่งมีหลักการพื้นฐานคือการแบ่งข้อมูลทุกหน่วยออกเป็นกลุ่มย่อย โดยผลรวมของระยะห่างระหว่างข้อมูลภายในกลุ่มกับจุดศูนย์กลางของกลุ่มนั้นมีค่าน้อยที่สุด โดยจุดศูนย์กลางกลุ่มคือข้อมูลใด ๆ ในชุดข้อมูลเท่านั้น ซึ่งเป็นหลักการเดียวกับอัลกอริทึมจัดกลุ่มแบบเคมีน (K-Means clustering algorithm) (Madhulatha 2011) แต่เนื่องจากอัลกอริทึมการจัดกลุ่มแบบเคมีนจะกำหนดค่าของจุดศูนย์กลางกลุ่มใหม่ ซึ่งคำนวณจากค่าเฉลี่ยของข้อมูลภายในกลุ่มนั้น ดังนั้นอัลกอริทึมการจัดกลุ่มแบบเคมีนจึงเหมาะสำหรับข้อมูลเชิงปริมาณเท่านั้น และไม่สามารถใช้ได้กับข้อมูลแบบผสม เพราะไม่

สามารถหาค่าเฉลี่ยของตัวแปรนามบัญญัติได้ ขณะที่อัลกอริทึมจัดกลุ่มแบบเคมีดอยด์ใช้ข้อมูลในชุดข้อมูลเท่านั้นเป็นจุดศูนย์กลางกลุ่ม จึงไม่มีปัญหาในสร้างจุดศูนย์กลางกลุ่มใหม่ เมทริกซ์ระยะห่างของข้อมูลจะคงที่ จึงสามารถปรับใช้การวิเคราะห์กลุ่มนี้กับเมทริกซ์ระยะห่างจากมาตรวัดความต่างแบบต่าง ๆ ได้

ผู้วิจัยจึงมีความสนใจว่า สำหรับข้อมูลแบบผสมที่ประกอบไปด้วยตัวแปรทั้งสามชนิดนี้ ระยะห่างแบบใดที่ทำให้การวิเคราะห์กลุ่มด้วยอัลกอริทึมจัดกลุ่มโดยรอบมีดอยด์มีประสิทธิภาพมากที่สุด โดยจำลองข้อมูลแบบผสมที่กำหนดให้ทราบกลุ่มที่แน่ชัดจากการกำหนดค่าพารามิเตอร์ของข้อมูลแต่ละกลุ่มแตกต่างกัน ประกอบกับการพิจารณาข้อมูลจริงที่เป็นข้อมูลแบบผสมและทราบกลุ่มของข้อมูลแน่ชัด ทั้งนี้นอกจากระยะห่างของ Kaufman and Rousseeuw (KR) และระยะห่างของ Podani (P) แล้ว ผู้วิจัยยังเสนอให้พิจารณาระยะห่างสำหรับตัวแปรนามบัญญัติของ Noorbehbahani et al. (N) ร่วมอีกด้วย นั่นคือระยะห่างของ Kaufman and Rousseeuw ร่วมกับ Noorbehbahani et al. (KR&N) และระยะห่างของ Podani ร่วมกับ Noorbehbahani et al. (P&N) ดังแสดงในตารางที่ 1.2

ตารางที่ 1.2 ระยะห่างสำหรับข้อมูลแบบผสมที่ใช้ระยะห่างสำหรับตัวแปรแต่ละประเภทแตกต่างกัน

ระยะห่างสำหรับข้อมูลแบบผสม	ระยะห่างสำหรับตัวแปรแต่ละประเภท		
	ตัวแปร เชิงปริมาณ	ตัวแปร นามบัญญัติ	ตัวแปรอันดับ
Kaufman and Rousseeuw (KR)	Gower	Gower	Kaufman and Rousseeuw
Podani (P)	Gower	Gower	Podani
Kaufman and Rousseeuw ร่วมกับ Noorbehbahani et al. (KR&N)	Gower	Noorbehbahani et al.	Kaufman and Rousseeuw
Podani ร่วมกับ Noorbehbahani et al. (P&N)	Gower	Noorbehbahani et al.	Podani

1.2 วัตถุประสงค์ของการวิจัย

1. เพื่อเสนอมาตรวัดระยะห่างสำหรับข้อมูลแบบผสม โดยประยุกต์ใช้ระยะห่างสำหรับตัวแปรนามบัญญัติของ Noorbehbahani et al. ร่วมกับระยะห่างของ Kaufman and Rousseeuw และระยะห่างของ Podani
2. เพื่อเปรียบเทียบประสิทธิภาพในการวิเคราะห์กลุ่มด้วยอัลกอริทึมจัดกลุ่มโดยรอบมีดอยด์ที่ใช้ระยะห่างของ Kaufman and Rousseeuw (KR) ระยะห่างของ Podani (P) ระยะห่างแบบ Kaufman and Rousseeuw ร่วมกับ Noorbehbahani et al. (KR&N) และระยะห่างแบบ Podani ร่วมกับ Noorbehbahani et al. (P&N) สำหรับข้อมูลแบบผสมที่ประกอบไปด้วยตัวแปรเชิงปริมาณ ตัวแปรนามบัญญัติ และตัวแปรอันดับ

1.3 ขอบเขตของการวิจัย

การวิจัยครั้งนี้เป็นการศึกษาเปรียบเทียบเทคนิคการวิเคราะห์กลุ่มข้อมูลแบบผสม โดยใช้มาตรวัดระยะห่างที่แตกต่างกัน ซึ่งมีขอบเขตการศึกษาดังต่อไปนี้

1. ตัวแปรที่ศึกษา
   1.1 ตัวแปรอิสระ (Independent variables; $\boldsymbol{X}$) จำนวน p ตัว นั่นคือ
      1.1.1 ตัวแปรนามบัญญัติ จำนวน p_{nom} ตัว
      โดยตัวแปรที่ l มีจำนวนประเภทเท่ากับ q_l ประเภท
      1.1.2 ตัวแปรอันดับ จำนวน p_{ord} ตัว
      โดยตัวแปรที่ l มีจำนวนอันดับเท่ากับ q_l อันดับ
      1.1.3 ตัวแปรเชิงปริมาณ จำนวน p_{quan} ตัว
      ดังนั้น $p = p_{nom} + p_{ord} + p_{quan}$ โดยที่ทุกตัวแปรแปลงมาจากตัวแปรเชิงปริมาณ $\boldsymbol{Z}$ ที่มีการแจกแจงแบบปกติหลายตัวแปร (Multivariate Normal Distribution) จำนวน p ตัว จากนั้นแปลงตัวแปรเชิงปริมาณเป็นตัวแปรนามบัญญัติและตัวแปรอันดับ โดยกำหนดค่าความน่าจะเป็นที่จะมีประเภทหรืออันดับต่าง ๆ สำหรับตัวแปรนามบัญญัติหรือตัวแปรอันดับที่ l
   1.2 ตัวแปรตาม (Dependent variables; Y) เป็นตัวแปรนามบัญญัติ จำนวน 1 ตัว และมีค่า $k = 1, 2, ..., K$ ซึ่ง K คือจำนวนกลุ่มของข้อมูล
2. กำหนดขนาดข้อมูลต่อกลุ่ม (n) เท่ากับ 20 และ 100 แทนกลุ่มข้อมูลขนาดเล็กและขนาดใหญ่ ตามลำดับ และกำหนดจำนวนกลุ่ม (K) เท่ากับ 3 และ 5 กลุ่ม ดังนั้นขนาดตัวอย่าง ($N = nK$) ที่เป็นไปได้ คือ 60 100 300 และ 500

3. กำหนดจำนวนตัวแปรอิสระ
   3.1 จำนวนตัวแปรนามบัญญัติ p_{nom} = 0 และ 3
   3.2 จำนวนตัวแปรอันดับ p_{ord} = 0 และ 3
   3.3 จำนวนตัวแปรเชิงปริมาณ p_{quan} = 0 และ 3
4. กำหนดจำนวนประเภทหรืออันดับ (q_l) สำหรับตัวแปรนามบัญญัติหรือตัวแปรอันดับที่ l เท่ากับ 5 ดังนั้น ตัวแปรนามบัญญัติหรือตัวแปรอันดับ แทนด้วย 1 2 3 4 และ 5 ซึ่งกำหนดให้ความน่าจะเป็นที่ประเภทหรืออันดับต่าง ๆ ของแต่ละตัวแปร เกิดขึ้นแตกต่างกันในแต่ละกลุ่ม เพื่อให้ข้อมูลตัวแปรนามบัญญัติหรือตัวแปรอันดับมีความแตกต่างระหว่างกลุ่ม
5. กำหนดค่าเฉลี่ยของตัวแปรเชิงปริมาณ แตกต่างกันในแต่ละกลุ่ม เพื่อให้ข้อมูลมีความแตกต่างระหว่างกลุ่ม
6. กำหนดความแปรปรวน

$$\sigma^2(Z_1) = \sigma^2(Z_2) = \ldots = \sigma^2(Z_p) = 1$$

7. กำหนดค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรอิสระ 2 กรณี ได้แก่
   7.1 ตัวแปรมีความสัมพันธ์ไปในทิศทางเดียวกัน น้อย (ρ = 0.2)
   7.2 ตัวแปรมีความสัมพันธ์ไปในทิศทางเดียวกัน มาก (ρ = 0.8)
8. กำหนดการจำลองข้อมูล 1,000 ครั้ง ต่อ 1 กรณี
9. มาตรวัดระยะห่างที่ใช้ในการศึกษาเป็นระยะห่างสำหรับข้อมูลแบบผสม ซึ่งมี 4 วิธีดังนี้
   9.1 ระยะห่างของ Kaufman and Rousseeuw (KR)
   9.2 ระยะห่างของ Podani (P)
   9.3 ระยะห่างแบบ Kaufman and Rousseeuw ร่วมกับ Noorbehbahani et al. (KR&N)
   9.4 ระยะห่างแบบ Podani ร่วมกับ Noorbehbahani et al. (P&N)
10. อัลกอริทึมที่ใช้ในการศึกษาการวิเคราะห์กลุ่มสำหรับข้อมูลแบบผสม คือ อัลกอริทึมจัดกลุ่มโดยรอบมีดอยด์ ด้วยฟังก์ชัน pam แพ็กเกจ cluster ในโปรแกรม R เวอร์ชัน 3.2.3
11. กำหนดข้อมูลจริง ได้แก่ ข้อมูล **Pittsburgh Bridges Data Set** จากเว็บไซต์ https://archive.ics.uci.edu/ml/datasets/Pittsburgh+Bridges
12. กำหนดค่าวัดประสิทธิภาพการวิเคราะห์กลุ่ม ดังนี้
   12.1 ค่า Purity
   12.2 ค่า Rand statistic
   12.3 ค่า Jaccard coefficient
   12.4 ค่า Average silhouette width

13. กำหนดการทดสอบความแตกต่างของค่าเฉลี่ยของค่าวัดประสิทธิภาพการวิเคราะห์กลุ่ม ดังนี้
    13.1 การวิเคราะห์ความแปรปรวนแบบมีปัจจัยเดียว (1-way ANOVA)
    13.2 การทดสอบความแตกต่างระหว่างค่าเฉลี่ยแต่ละคู่ของค่าที่วัดได้ ด้วยวิธี Tukey
    13.3 การทดสอบความแตกต่างระหว่างค่าเฉลี่ยแต่ละคู่ของค่าที่วัดได้ ด้วยสถิติทดสอบ t

1.4 คำจำกัดความที่ใช้เฉพาะการวิจัย

1. ตัวแปรเชิงปริมาณ (Quantitative variable)

   หมายถึง ตัวแปรที่มีค่ากำหนดเป็นตัวเลขได้ และสามารถบอกได้ว่าระยะห่างระหว่างกันมากน้อยเพียงใด เช่น น้ำหนัก ส่วนสูง อุณหภูมิ คะแนนสอบ เป็นต้น

2. ตัวแปรนามบัญญัติ (Nominal variable)

   หมายถึง ตัวแปรที่จำแนกสิ่งต่าง ๆ ออกเป็นกลุ่ม ไม่สามารถกำหนดค่าเป็นตัวเลขได้แน่นอน และไม่สามารถเปรียบเทียบว่าสิ่งนั้น ๆ มีค่ามากกว่า น้อยกว่า หรือดีกว่ากัน เช่น อาชีพ สถานภาพสมรส เป็นต้น

3. ตัวแปรอันดับ (Ordinal variable)

   หมายถึง ตัวแปรที่ให้ข้อมูลแบบจัดอันดับ ให้ค่าเป็นตัวเลข สามารถเปรียบเทียบว่าสิ่งนั้น ๆ มีค่ามากกว่า น้อยกว่า หรือดีกว่า แต่ไม่สามารถบอกค่าความต่างที่แท้จริงได้ เช่น ระดับการศึกษา ระดับความคิดเห็น เป็นต้น

4. ข้อมูลแบบผสม (Mixed data)

   หมายถึง ข้อมูลที่ประกอบไปด้วยตัวแปรเชิงปริมาณ ตัวแปรนามบัญญัติ และตัวแปรอันดับ

5. คะแนนอันดับ (Rank score)

   หมายถึง ค่าที่ได้จากการแปลงข้อมูลตัวแปรอันดับ โดยการเรียงข้อมูลทั้งหมดของตัวแปรนั้นจากน้อยไปมาก ให้มีอันดับตั้งแต่ 1 ถึงจำนวนข้อมูลทั้งหมด จากนั้นบวกรวมค่าอันดับใหม่ที่เดิมมีอันดับเดียวกันทั้งหมด แล้วจึงหารด้วยจำนวนการซ้ำกันของอันดับนั้น ๆ เช่น ชุดข้อมูล 1, 1, 1, 2, 3 และ 3 มีข้อมูลทั้งสิ้น 6 ตัว เรียงเป็นข้อมูลชุดใหม่คือ 1, 2, 3, 4, 5 และ 6 เมื่อพิจารณาการซ้ำกันสามารถแปลงได้เป็นค่าอันดับใหม่ คือ 2, 2, 2, 4, 5.5 และ 5.5

6. ระยะห่างของ Kaufman and Rousseeuw (KR)

   หมายถึง ระยะห่างสำหรับข้อมูลแบบผสม โดยวัดระยะห่างข้อมูลที่เป็นตัวแปรเชิงปริมาณและตัวแปรนามบัญญัติตามนิยามของ Gower และวัดระยะห่างข้อมูลที่เป็นตัวแปรอันดับตามนิยามของ Kaufman and Rousseeuw

7. ระยะห่างของ Podani (P)

   หมายถึง ระยะห่างสำหรับข้อมูลแบบผสม โดยวัดระยะห่างข้อมูลที่เป็นตัวแปรเชิงปริมาณและตัวแปรนามบัญญัติตามนิยามของ Gower และวัดระยะห่างข้อมูลที่เป็นตัวแปรอันดับตามนิยามของ Podani

8. ระยะห่างแบบ Kaufman and Rousseeuw ร่วมกับ Noorbehbahani et al. (KR&N)

   หมายถึง ระยะห่างสำหรับข้อมูลแบบผสม โดยวัดระยะห่างตัวแปรข้อมูลที่เป็นตัวแปรเชิงปริมาณตามนิยามของ Gower วัดระยะห่างข้อมูลที่เป็นตัวแปรนามบัญญัติตามนิยามของ Noorbehbahani et al. และวัดระยะห่างข้อมูลที่เป็นตัวแปรอันดับตามนิยามของ Kaufman and Rousseeuw

9. ระยะห่างแบบ Podani ร่วมกับ Noorbehbahani et al. (P&N)

   หมายถึง ระยะห่างสำหรับข้อมูลแบบผสม โดยวัดระยะห่างข้อมูลที่เป็นตัวแปรเชิงปริมาณตามนิยามของ Gower วัดระยะห่างข้อมูลที่เป็นตัวแปรนามบัญญัติตามนิยามของ Noorbehbahani et al. และวัดระยะห่างข้อมูลที่เป็นตัวแปรอันดับตามนิยามของ Podani

1.5 วิธีดำเนินการวิจัย

1. ศึกษามาตรวัดระยะห่างของข้อมูลแบบผสม
2. กำหนดค่าเริ่มต้นสำหรับจำลองข้อมูลแบบผสม
3. ทำการจำลองข้อมูลแบบผสมตามลักษณะที่กำหนดไว้ข้างต้น จำนวน 1,000 ครั้ง
4. นำข้อมูลที่จำลองขึ้นมาหาเมทริกซ์ระยะห่างด้วยมาตรวัดระยะห่าง ดังนี้
   4.1 ระยะห่างของ Kaufman and Rousseeuw (KR)
   4.2 ระยะห่างของ Podani (P)
   4.3 ระยะห่างแบบ Kaufman and Rousseeuw ร่วมกับ Noorbehbahani et al. (KR&N)
   4.4 ระยะห่างแบบ Podani ร่วมกับ Noorbehbahani et al. (P&N)
5. วิเคราะห์กลุ่มข้อมูลที่จำลองขึ้นและตัวอย่างข้อมูลจริง ด้วยอัลกอริทึมจัดกลุ่มโดยรอบมีดอยด์ โดยใช้เมทริกซ์ระยะห่างจากขั้นตอนที่ 5

6. คำนวณค่า Purity ค่า Rand statistic ค่า Jaccard coefficient เพื่อเปรียบเทียบประสิทธิภาพการวิเคราะห์กลุ่ม และค่า Average silhouette width
7. เปรียบเทียบประสิทธิภาพของระยะห่างแบบต่าง ๆ โดยคำนวณค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน กราฟช่วงความเชื่อมั่น 95% วิเคราะห์ความแปรปรวน และทดสอบความแตกต่างของค่าเฉลี่ย จากผลการวิเคราะห์กลุ่มในขั้นตอนที่ 6
8. กำหนดตัวอย่างข้อมูลจริงที่จะใช้ในการศึกษา
9. วิเคราะห์กลุ่มตัวอย่างข้อมูลจริงด้วยระยะห่างทั้ง 4 วิธี และศึกษาผลการวิเคราะห์กลุ่มตัวอย่างข้อมูลจริง

1.6 ประโยชน์ที่คาดว่าจะได้รับ

เพื่อเป็นแนวทางในการเลือกใช้ระยะห่างสำหรับการวิเคราะห์กลุ่มได้อย่างเหมาะสมกับลักษณะของข้อมูลแบบผสม

บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 มาตรวัดระยะห่างสำหรับข้อมูลแบบผสม

Gower (1971) ได้เสนอมาตรวัดความคล้าย สำหรับข้อมูลที่ประกอบไปด้วยตัวแปรเชิงปริมาณและตัวแปรนามบัญญัติ อย่างไรก็ตาม การวิเคราะห์กลุ่มส่วนใหญ่ไม่นิยมวัดความคล้ายระหว่างข้อมูลในการวิเคราะห์ มาตรวัดความคล้ายของ Gower จึงถูกปรับให้อยู่ในรูปของมาตรวัดระยะห่าง นั่นคือ ระยะห่างระหว่างข้อมูลที่ i กับ j

$$D_{ij} = \frac{\sum_{l=1}^{p} d_{ijl}\delta_{ijl}}{\sum_{l=1}^{p} \delta_{ijl}}$$

โดยที่ p แทน จำนวนตัวแปรของชุดข้อมูล

δ_{ijl} เท่ากับ 0 เมื่อไม่ทราบค่าของข้อมูลที่ i หรือ j ของตัวแปรที่ l

δ_{ijl} เท่ากับ 1 เมื่อทราบค่าของข้อมูลที่ i หรือ j ของตัวแปรที่ l

และ d_{ijl} แทน ระยะห่างระหว่างข้อมูลที่ i กับ j วัดโดยตัวแปรที่ l

จะเห็นว่า D_{ij} คือระยะห่างระหว่างข้อมูล โดยที่มีค่าขึ้นอยู่กับ d_{ijl} ซึ่งเป็นระยะห่างที่ถูกวัดโดยตัวแปร ดังนั้น d_{ijl} จึงถูกคำนวณด้วยวิธีที่แตกต่างกันขึ้นอยู่กับประเภทของตัวแปรที่ l นั้น ๆ โดย Gower ได้กล่าวถึงวิธีการวัดระยะห่างสำหรับตัวแปรเชิงปริมาณ และตัวแปรนามบัญญัติ ดังนี้

2.1.1 ระยะห่างสำหรับตัวแปรเชิงปริมาณของ Gower

$$d_{ijl} = \frac{\left| x_{il} - x_{jl} \right|}{R_l}$$

โดยที่ x_{hl} แทน ค่าของข้อมูลที่ h ของตัวแปรเชิงปริมาณที่ l เมื่อ $h = i, j$

และ R_l แทน พิสัยของข้อมูลตัวแปรตัวที่ l

2.1.2 ระยะห่างสำหรับตัวแปรนามบัญญัติของ Gower

d_{ijl} เท่ากับ 0 ถ้าข้อมูลที่ i กับ j ของตัวแปรนามบัญญัติที่ l อยู่ต่างประเภทกัน

d_{ijl} เท่ากับ 1 ถ้าข้อมูลที่ i กับ j ของตัวแปรนามบัญญัติที่ l อยู่ในประเภทเดียวกัน

ตัวอย่างการคำนวณระยะห่างสำหรับตัวแปรเชิงปริมาณและตัวแปรนามบัญญัติของ Gower

ข้อมูลลูกค้าจำนวน 6 คนในตารางที่ 2.1 เป็นข้อมูลรายได้ (บาท) และอาชีพ โดยให้ 1 แทนอาชีพข้าราชการ 2 แทนอาชีค้าขาย และ 3 แทนอาชีพรับจ้า/ลูกจ้าง ตัวแปรนามบัญญัติของข้อมูลชุดนี้จึงมีจำนวนประเภท เท่ากับ 3

ตารางที่ 2.1 ตัวอย่างข้อมูลที่ประกอบด้วยตัวแปรเชิงปริมาณและตัวแปรนามบัญญัติ อย่างละ 1 ตัวแปร

ลูกค้าที่		1	2	3	4	5	6
ตัวแปรที่ 1	รายได้ (บาท)	450	300	500	300	800	200
ตัวแปรที่ 2	อาชีพ	1	3	2	3	2	2

ระยะห่างระหว่างลูกค้าที่ 1 กับ 2 วัดโดยตัวแปรที่ 1 รายได้ (ตัวแปรเชิงปริมาณ) เท่ากับ

$$d_{121} = \frac{|x_{11} - x_{21}|}{R_1} = \frac{|450 - 300|}{800 - 200} = \frac{150}{600} = 0.25$$

ระยะห่างระหว่างลูกค้าที่ 1 กับ 2 วัดโดยตัวแปรที่ 2 อาชีพ (ตัวแปรนามบัญญัติ) เท่ากับ

$$d_{122} = 1$$

คำนวณ ระยะห่างระหว่างลูกค้าคนที่ 1 กับ 2 ได้ดังนี้

$$D_{12} = \frac{\sum_{l=1}^{2} d_{ijl}\delta_{ijl}}{\sum_{l=1}^{2} \delta_{ijl}} = \frac{(0.25 \times 1) + (1 \times 1)}{1+1} = 0.625$$

อย่างไรก็ตาม Gower ไม่ได้กล่าวถึงความคล้ายสำหรับตัวแปรอันดับ ซึ่งอาจเป็นตัวแปรสำคัญในการวิเคราะห์กลุ่ม Kaufman and Rousseeuw (1990) จึงเสนอมาตรวัดความระยะห่างสำหรับตัวแปรอันดับเพิ่มเติมจากระยะห่างของ Gower เพื่อให้สามารถวัดระยะห่างข้อมูลแบบผสมได้ โดยแปลงข้อมูลที่เป็นตัวแปรอันดับให้มีค่าอยู่ระหว่าง 0 ถึง 1

2.1.3 ระยะห่างสำหรับตัวแปรอันดับของ Kaufman and Rousseeuw

$$d_{ijl} = \frac{\left| z_{il} - z_{jl} \right|}{R_l}$$

โดยที่

$$z_{hl} = \frac{x_{hl} - 1}{M_l - 1}$$

เมื่อ x_{hl} แทน อันดับของข้อมูลที่ h ของตัวแปรอันดับที่ l สำหรับ $h = i, j$

M_l แทน ค่าอันดับสูงสุดของตัวแปรอันดับที่ l

และ R_l แทน พิสัยของข้อมูลตัวแปรอันดับที่ l ที่ถูกแปลงค่าเป็น z_{hl} แล้ว

นอกจากนี้ Podani (1999) ได้สร้างมาตรวัดความคล้ายสำหรับตัวแปรอันดับเพิ่มเติมจากวิธีเดิมของ Gower เช่นกัน โดยแปลงข้อมูลตัวแปรอันดับเป็นคะแนนอันดับ (Rank Score) ซึ่งเป็นค่าที่ได้จากการเรียงข้อมูลทั้งหมดของตัวแปรนั้นจากน้อยไปมาก และกำหนดอันดับใหม่ตั้งแต่ 1 ถึงจำนวนข้อมูลทั้งหมด จากนั้นบวกรวมค่าอันดับใหม่ที่เดิมมีอันดับเดียวกันทั้งหมด แล้วจึงหารด้วยจำนวนการซ้ำกันของอันดับนั้น ๆ เช่น ชุดข้อมูล 1, 1, 1, 2, 3 และ 3 มีข้อมูลทั้งสิ้น 6 ตัว เรียงเป็นข้อมูลชุดใหม่ คือ 1, 2, 3, 4, 5 และ 6 เมื่อพิจารณาการซ้ำกันสามารถแปลงได้เป็นค่าอันดับใหม่ คือ 2, 2, 2, 4, 5.5 และ 5.5 กำหนดให้ใช้คะแนนอันดับแทนค่าอันดับเริ่มต้นและนำไปคำนวณด้วยความคล้ายซึ่ง Podani ได้สร้างขึ้นใหม่ 2 วิธี สำหรับวิธีแรก เมื่อแปลงเป็นค่าความต่างแล้ว จะได้เมทริกซ์ระยะห่าง (Distance matrix) ของข้อมูลที่มีคุณสมบัติเป็นเมทริก (Metric) ขณะที่วิธีที่สอง จะได้เมทริกซ์ระยะห่างที่ไม่มีคุณสมบัติเป็นเมทริก (Non-metric) วิธีนี้จึงไม่สามารถนำไปใช้กับเทคนิคการวิเคราะห์กลุ่มที่ต้องใช้เมทริกซ์ระยะห่างที่มีคุณสมบัติเป็นเมทริกได้ ทั้งนี้สนใจศึกษามาตรวัดของ Podani ที่เมื่อพิจารณาระยะห่างของข้อมูลทุกคู่แล้วได้เป็นเมทริกซ์ระยะห่าง ตามนิยามดังต่อไปนี้

2.1.4 ระยะห่างสำหรับตัวแปรอันดับของ Podani

$$d_{ijl} = \frac{\left| r_{il} - r_{jl} \right|}{R_l}$$

โดยที่ r_{hl} แทน คะแนนอันดับของข้อมูลที่ h ของตัวแปรอันดับที่ l เมื่อ $h = i, j$
และ R_l แทน พิสัยของข้อมูลตัวแปรอันดับที่ l ที่ถูกแปลงเป็นคะแนนอันดับแล้ว

ตัวอย่างการคำนวณระยะห่างสำหรับตัวแปรอันดับของ Kaufman and Rousseeuw และ Podani

ข้อมูลลูกค้าจำนวน 6 คน ดังแสดงในตารางที่ 2.2 เป็นข้อมูลระดับการศึกษา โดยให้ 1 แทนระดับชั้นมัธยม 2 แทนระดับปริญญาตรี และ 3 แทนระดับปริญญาโท ตัวแปรอันดับของข้อมูลชุดนี้จึงมีอันดับเท่ากับ 3

ตารางที่ 2.2 ตัวอย่างข้อมูลที่เป็นตัวแปรอันดับ

ลูกค้าที่		1	2	3	4	5	6	R_1
ตัวแปรที่ 1	ระดับการศึกษา	2	1	3	1	2	2	-
	z_{i1} (Kaufman and Rousseeuw)	0.5	0	1	0	0.5	0.5	1
	r_{i1} (คะแนนอันดับ – Podani)	4	1.5	6	1.5	4	4	4.5

- ระยะห่างสำหรับตัวแปรอันดับของ Kaufman and Rousseeuw

  ระยะห่างระหว่างลูกค้าที่ 1 กับ 2 วัดโดยตัวแปรที่ 1 เท่ากับ

$$d_{121} = \frac{|z_{11} - z_{21}|}{R_1} = \frac{|0.5 - 0|}{1} = 0.5$$

- ระยะห่างสำหรับตัวแปรอันดับของ Podani

  ระยะห่างระหว่างลูกค้าที่ 1 กับ 2 วัดโดยตัวแปรที่ 1 เท่ากับ

$$d_{121} = \frac{|r_{11} - r_{21}|}{R_1} = \frac{|4 - 1.5|}{4.5} = 0.556$$

Noorbehbahani et al. (2015) ได้นำเสนอมาตรวัดระยะห่างสำหรับข้อมูลที่ประกอบไปด้วยตัวแปรเชิงปริมาณและตัวแปรนามบัญญัติ ซึ่งแตกต่างจากวิธีของ Gower ทั้งนี้ Noorbehbahani et al. เล็งเห็นว่ามาตรวัดระยะห่างของตัวแปรนามบัญญัติไม่ควรมีค่าเพียง 0 และ 1 เท่านั้น แต่ควรมีค่าที่ขึ้นอยู่กับความถี่ของค่าของตัวแปรนั้น ๆ กล่าวคือสำหรับค่าที่ต่างกันของตัวแปรนามบัญญัติ ค่าของตัวแปรที่มีความถี่สูง (เกิดขึ้นบ่อย) สองค่าควรมีระยะห่างน้อย ขณะที่ค่าของตัวแปรที่มีความถี่ต่ำสองค่า หรือค่าของตัวแปรสองค่าที่ค่าหนึ่งมีความถี่ต่ำกับอีกค่ามีความถี่สูง ควรมีระยะห่างที่สูงขึ้น เนื่องจากโอกาสที่จะถูกจัดให้อยู่ในกลุ่มเดียวกันไม่มากนัก อย่างไรก็ตาม Noorbehbahani et al. พิจารณาเพียงกรณีที่ข้อมูลประกอบไปด้วยตัวแปรเชิงปริมาณและตัวแปรนามบัญญัติเท่านั้น และยังให้ค่าน้ำหนักในการวัดระยะห่างสำหรับตัวแปรสองประเภทนี้ไม่เท่ากันอีกด้วย แต่ผู้วิจัยเล็งเห็นว่าระยะห่างสำหรับตัวแปรนามบัญญัติของ Noorbehbahani et al. ที่มีค่าขึ้นกับจำนวนความถี่ของข้อมูลสามารถนำไปประยุกต์ใช้ร่วมกับระยะห่างของ Gower ได้ และจะทำให้ประสิทธิภาพในการวัดระยะห่างสำหรับข้อมูลแบบผสมแม่นยำขึ้น

2.1.5 ระยะห่างสำหรับตัวแปรนามบัญญัติของ Noorbehbahani et al.

ถ้าข้อมูลที่ i กับ j ของตัวแปรนามบัญญัติที่ l อยู่ในประเภทเดียวกัน กำหนดให้

$$d_{ijl} = 0$$

ถ้าข้อมูลที่ i กับ j ของตัวแปรนามบัญญัติที่ l อยู่ต่างประเภทกัน

$$d_{ijl} = \frac{|f_{il} - f_{jl}| + \min\{f_l\}}{\max\{f_{il}, f_{jl}\}}$$

โดยที่ f_{hl} แทน ความถี่ของข้อมูลที่มีประเภทเดียวกันกับข้อมูลตัวที่ h ของตัวแปรนามบัญญัติที่ l เมื่อ $h = i, j$

$\min\{f_l\}$ แทน ความถี่ที่น้อยที่สุดจากประเภทข้อมูลทั้งหมดของตัวแปรนามบัญญัติที่ l

และ $\max\{f_{il}, f_{jl}\}$ แทน ค่าที่มากที่สุดระหว่าง f_{il} และ f_{jl}

ตัวอย่างการคำนวณระยะห่างสำหรับตัวแปรนามบัญญัติของ Noorbehbahani et al.

ข้อมูลลูกค้าจำนวน 6 คน เป็นข้อมูลอาชีพ มีจำนวนประเภทเท่ากับ 3 โดย 1 แทนอาชีพข้าราชการ 2 แทนอาชีพค้าขาย และ 3 แทนอาชีพรับจ้าง/ลูกจ้าง ดังแสดงในตารางที่ 2.3

ตารางที่ 2.3 ตัวอย่างข้อมูลตัวแปรนามบัญญัติจำนวน 1 ตัวแปร

ลูกค้าที่		1	2	3	4	5	6
ตัวแปรที่ 1	อาชีพ	1	3	3	3	2	2
	f_{i1}	1	3	3	3	2	2

จากข้อมูลจะเห็นว่า ลูกค้าที่ประกอบอาชีพข้าราชการ ค้าขาย และรับจ้าง/ลูกจ้าง มีจำนวน 1 2 และ 3 คน ตามลำดับ จึงได้ค่าความถี่ f_{i1} ดังแสดงในตารางที่ 2.3 และความถี่ที่น้อยที่สุดของอาชีพทั้งสามนี้ หรือ $\min\{f_1\}$ คือ 1

ระยะห่างระหว่างลูกค้าที่ 1 กับ 2 วัดโดยตัวแปรที่ 1 เท่ากับ

$$d_{ijl} = \frac{|f_{11} - f_{21}| + \min\{f_1\}}{\max\{f_{11}, f_{21}\}} = \frac{|1-3|+1}{\max\{1,3\}} = \frac{3}{3} = 1$$

2.2 อัลกอริทึมจัดกลุ่มโดยรอบมีดอยด์

อัลกอริทึมจัดกลุ่มโดยรอบมีดอยด์ (Partitioning around medoids algorithm, PAM algorithm) เป็นอัลกอริทึมจัดกลุ่มรูปแบบหนึ่งของอัลกอริทึมจัดกลุ่มแบบเคมีดอยด์ (K-Medoids clustering algorithm) ซึ่งมีหลักการพื้นฐานคือการแบ่งข้อมูลทุกหน่วยออกเป็นกลุ่มย่อย โดยผลรวมของระยะห่างระหว่างข้อมูลภายในกลุ่มกับจุดศูนย์กลาง หรือมีดอยด์ (Medoid) ของกลุ่มนั้นมีค่าน้อยที่สุด ซึ่งหลักการนี้คล้ายคลึงกับอัลกอริทึมจัดกลุ่มแบบเคมีน (K-Means clustering algorithm) (Madhulatha, 2011: 477)

ความแตกต่างระหว่างอัลกอริทึมจัดกลุ่มแบบเคมีดอยด์และแบบเคมีน คือ อัลกอริทึมจัดกลุ่มแบบเคมีดอยด์จะเลือกข้อมูลในชุดข้อมูลเท่านั้นเป็นจุดศูนย์กลาง ขณะที่อัลกอริทึมการจัดกลุ่มแบบเคมีนจะกำหนดค่าของจุดศูนย์กลางใหม่ ซึ่งเรียกว่าเซ็นทรอยด์ (Centroid) โดยคำนวณจากค่าเฉลี่ยของข้อมูลภายในกลุ่มนั้น แสดงว่าอัลกอริทึมจัดกลุ่มแบบเคมีนใช้ได้เฉพาะข้อมูลที่เป็นตัวแปรเชิงปริมาณเท่านั้น ขณะที่อัลกอริทึมจัดกลุ่มแบบเคมีดอยด์จะพิจารณาเพียงข้อมูลในชุดข้อมูล ทำให้ระยะห่างระหว่างข้อมูลแต่ละคู่ไม่เปลี่ยนแปลงตลอดการวิเคราะห์กลุ่ม จึงสามารถทำการวิเคราะห์กลุ่มสำหรับข้อมูลแบบผสมด้วยระยะห่างแบบต่าง ๆ ด้วยอัลกอริทึมการจัดกลุ่มแบบเคมีดอยด์ได้

อัลกอริทึมจัดกลุ่มโดยรอบมีดอยด์ประกอบไปด้วย 2 ระยะ ระยะแรกเรียกว่า BUILD เพื่อหาตัวแทนข้อมูล (Representative object) ทั้งหมด มีจำนวนเท่ากับจำนวนกลุ่มที่ต้องการ (K) ซึ่งตัวแทนข้อมูลตัวแรกคือข้อมูลที่ให้ผลรวมระยะห่างระหว่างข้อมูลตัวแรกกับข้อมูลทั้งหมดน้อยที่สุด จากนั้นจึงหาตัวแทนข้อมูลตัวถัดไปที่ให้ผลรวมระยะห่างมีค่ารองลงมา ทำซ้ำจนกระทั่งได้ตัวแทนข้อมูลครบตามจำนวนกลุ่มที่ต้องการ ระยะที่สองเรียกกว่า SWAP ทำการสับเปลี่ยนตัวแทนข้อมูลกับข้อมูลที่ไม่ถูกเลือกให้เป็นตัวแทนข้อมูล มีวัตถุประสงค์ในการปรับปรุงตัวแทนข้อมูลทั้งหมด เพื่อหามีดอยด์ที่แท้จริงที่ให้ผลรวมระยะห่างระหว่างมีดอยด์นั้น ๆ กับข้อมูลภายในกลุ่มเดียวกันมีค่าน้อยที่สุด

กำหนดให้ I_i แทน ข้อมูลที่ i

ระยะที่ 1 BUILD

1. พิจารณา I_i ที่ยังไม่ถูกเลือกให้เป็นตัวแทนข้อมูล
2. พิจารณา I_j ที่ $j \neq i$ และคำนวณผลต่างระหว่าง ระยะห่างของ I_j กับตัวแทนข้อมูลที่ถูกเลือกไปก่อนหน้านี้ (D_j) และ ระยะห่างของ I_j กับ I_i ($d(j,i)$)
3. ถ้าผลต่าง $D_j - d(j,i)$ เป็นบวก แล้ว I_j จะถูกจัดให้อยู่ในกลุ่มเดียวกับ I_i และคำนวณ
$$C_{ji} = \max(D_j - d(j,i), 0)$$
4. คำนวณผลรวม $\sum_j C_{ji}$
5. เลือก I_i ที่ให้ค่า $\sum_j C_{ji}$ สูงที่สุด เป็นตัวแทนข้อมูล
6. ทำซ้ำขั้นตอนที่ 1. ถึง 5. จนกระทั่งได้ตัวแทนข้อมูล K ตัว

ระยะที่ 2 SWAP

พิจารณาการสับเปลี่ยน I_i ที่ถูกเลือกเป็นตัวแทนข้อมูล กับ I_h ที่ไม่ถูกเลือกเป็นตัวแทนข้อมูล

1. พิจารณา I_j ที่ไม่ถูกเลือกเป็นตัวแทนข้อมูล และคำนวณค่า C_{jih} ซึ่งเป็นค่าสำหรับการพิจารณาการสับเปลี่ยนระหว่าง I_i ที่ถูกเลือกเป็นตัวแทนข้อมูล กับ I_h ที่ไม่ถูกเลือกก่อนหน้านี้ โดยที่
   1.1 ถ้า I_j มีระยะห่างกับทั้ง I_i และ I_h มากกว่าหนึ่งในตัวแทนข้อมูลอื่น ๆ
   $$C_{jih} = 0$$
   1.2 ถ้า I_j มีระยะห่างกับ I_i น้อยกว่าตัวแทนข้อมูลอื่น ๆ ตัวใดตัวหนึ่ง ($d(j,i) = D_j$) ต้องพิจารณาสถานการณ์ต่อไปนี้
      1.1.1 I_j มีระยะห่างกับ I_h น้อยกว่าระยะห่างกับตัวแทนข้อมูลที่ให้ผลรวมระยะห่างระหว่างตัวแทนข้อมูลนั้นกับข้อมูลทั้งหมดน้อยที่สุดเป็นอันดับสอง นั่นคือ

$$d(j,h) < E_j$$

ซึ่ง E_j คือระยะห่างระหว่าง I_j กับตัวแทนข้อมูลที่ให้ผลรวมระยะห่างระหว่างตัวแทนข้อมูลนั้นกับข้อมูลทั้งหมดน้อยที่สุดเป็นอันดับสอง ในกรณีนี้

$$C_{jih} = d(j,h) - d(j,i)$$

1.1.2 I_j มีระยะห่างกับ I_h มากกว่าหรือเท่ากับระยะห่างกับตัวแทนข้อมูลที่ให้ผลรวมระยะห่างระหว่างตัวแทนข้อมูลนั้นกับข้อมูลทั้งหมดน้อยที่สุดเป็นอันดับสอง นั่นคือ

$$d(j,h) \geq E_j$$

จะได้ว่า

$$C_{jih} = E_j - D_j$$

1.3 ถ้า I_j มีระยะห่างกับ I_i มากกว่าระยะห่างกับตัวแทนข้อมูลอื่น ๆ อย่างน้อยหนึ่งตัว แต่ I_j มีระยะห่างกับ I_h น้อยกว่าระยะห่างกับตัวแทนข้อมูลอื่น ๆ จะได้ว่า

$$C_{jih} = d(j,h) - D_j$$

2. คำนวณผลรวม

$$T_{ih} = \sum_j C_{jih}$$

3. เลือก i และ h ที่

$$\underset{i,h}{minimizes} T_{ih}$$

4. ถ้า $T_{ih} < 0$ จะทำการสับเปลี่ยนตัวแทนข้อมูลระหว่าง I_i กับ I_h และทำซ้ำตั้งแต่ขั้นตอนที่ 1. ถึง 4. ในระยะที่ 2

ถ้า $T_{ih} \geq 0$ แสดงว่าผลรวมระยะห่างระหว่างตัวแทนข้อมูลนั้น ๆ กับข้อมูลภายในกลุ่มเดียวกันมีค่าน้อยที่สุด จึงได้ว่าตัวแทนข้อมูลที่ได้คือมีดอยด์ที่ต้องการ จึงสิ้นสุดกระบวนการ

2.3 การวัดประสิทธิภาพการวิเคราะห์กลุ่ม

สำหรับข้อมูลที่ทราบกลุ่มที่แท้จริง นั่นคือทราบอยู่แล้วว่าข้อมูลแต่ละตัวนั้นอยู่ในกลุ่มใด สามารถวัดประสิทธิภาพการวิเคราะห์กลุ่มด้วย ค่า Rand statistic ค่า Jaccard coefficient และค่า Purity ขณะที่ค่า Average silhouette width เป็นค่าที่วัดประสิทธิภาพการวิเคราะห์กลุ่ม ในกรณีที่ไม่ทราบกลุ่มที่แท้จริงของข้อมูล โดยเปรียบเทียบค่า Average silhouette width ที่ได้จากการวิเคราะห์กลุ่มที่กำหนดจำนวนกลุ่มแตกต่างกัน เพื่อหาจำนวนกลุ่มที่เหมาะสมที่สุดสำหรับข้อมูลชุดนั้น

กำหนดให้ I_i แทน ข้อมูลที่ i โดยที่ทราบว่า I_i อยู่ในกลุ่มที่ $L(I_i)$ และเมื่อทำการวิเคราะห์กลุ่ม พบว่า I_i ถูกจัดอยู่ในกลุ่มที่ $C(I_i)$

กำหนดให้ $SS = \{(I_i, I_j) | C(I_i) = C(I_j)$ และ $L(I_i) = L(I_j)\}$

$SD = \{(I_i, I_j) | C(I_i) = C(I_j)$ และ $L(I_i) \neq L(I_j)\}$

$DS = \{(I_i, I_j) | C(I_i) \neq C(I_j)$ และ $L(I_i) = L(I_j)\}$

$DD = \{(I_i, I_j) | C(I_i) \neq C(I_j)$ และ $L(I_i) \neq L(I_j)\}$

2.3.1 ค่า Rand statistic

$$R = \frac{|SS| + |DD|}{|SS| + |SD| + |DS| + |DD|}$$

ดังนั้น $0 \leq R \leq 1$ ซึ่งค่า Rand statistic จะพิจารณาการวิเคราะห์กลุ่มที่ให้ผลออกมาถูกต้องในภาพรวม คือพิจารณาทั้งกรณีข้อมูลแต่ละคู่ที่ทราบว่าอยู่กลุ่มเดียวกันและถูกจัดให้อยู่กลุ่มเดียวกัน และกรณีข้อมูลแต่ละคู่ควรอยู่ต่างกลุ่มกันและถูกจัดให้อยู่ต่างกลุ่มกัน (Rand 1971) ดังนั้นถ้า R มีค่าใกล้ 1 นั่นคือในภาพรวมการวิเคราะห์กลุ่มมีความถูกต้องมาก และถ้า R มีค่าใกล้ 0 นั่นคือในภาพรวมการวิเคราะห์กลุ่มมีความถูกต้องน้อย

2.3.2 ค่า Jaccard coefficient

$$J = \frac{|SS|}{|SS| + |SD| + |DS|}$$

ดังนั้น $0 \leq J \leq 1$ ซึ่งค่า Jaccard coefficient จะพิจารณาการวิเคราะห์กลุ่มที่ให้ผลออกมาถูกต้องเฉพาะกรณีข้อมูลแต่ละคู่ที่ทราบว่าอยู่กลุ่มเดียวกันและยังถูกจัดให้อยู่กลุ่มเดียวกัน (Manning, Raghavan et al. 2009) ดังนั้น ถ้า J มีค่าใกล้ 1 นั่นคือการวิเคราะห์กลุ่มให้ผลถูกต้องใกล้เคียงกับกลุ่มที่แท้จริง และถ้า J มีค่าใกล้ 0 นั่นคือการวิเคราะห์กลุ่มให้ผลถูกต้องไม่ใกล้เคียงกับกลุ่มที่แท้จริง

2.3.3 ค่า Purity

$$purity(L, C) = \frac{\sum_k \max_h |L_h \bigcap C_k|}{N}$$

เมื่อ $L = \{L_1, L_2, \ldots, L_H\}$ แทน เซตของกลุ่มของข้อมูล H กลุ่มที่ทราบกลุ่มที่แท้จริง

$C = \{C_1, C_2, \ldots, C_K\}$ แทน เซตของกลุ่มของข้อมูล K กลุ่มที่ได้จากการวิเคราะห์กลุ่ม

N แทน จำนวนข้อมูลทั้งหมด

ดังนั้น $0 \le purity(L,C) \le 1$ โดยค่า Purity จะพิจารณาส่วนที่ซ้อนทับกันของกลุ่มข้อมูลที่แท้จริงกับกลุ่มข้อมูลที่ได้จากการวิเคราะห์กลุ่ม ถ้าหากค่า Purity ใกล้ 1 แสดงว่าการวิเคราะห์กลุ่มมีความถูกต้องมาก ถ้าหากค่า Purity ใกล้ 0 แสดงว่าการวิเคราะห์กลุ่มมีความถูกต้องน้อย (Manning, Raghavan et al. 2009)

2.3.4 ค่า Average silhouette width

ค่า Average silhouette width คือค่าเฉลี่ยของค่า Silhouette จากข้อมูลทั้งหมดในการวิเคราะห์กลุ่ม เป็นค่าที่บอกจำนวนกลุ่มที่เหมาะสมจากการวิเคราะห์กลุ่ม เมื่อไม่ทราบจำนวนกลุ่มที่แท้จริงหรือจำนวนกลุ่มที่ต้องการ โดยเปรียบเทียบผลการวิเคราะห์กลุ่มที่กำหนดจำนวนกลุ่มที่แตกต่างกัน (Rousseeuw 1987) หากการวิเคราะห์กลุ่มที่มีจำนวนกลุ่มแตกต่างกันแบบใดมีค่า Average silhouette width มากกว่า แสดงว่ากลุ่มที่กำหนดนั้นมีความเหมาะสมกับข้อมูลมากกว่า

กำหนด ค่า Silhouette ของข้อมูลที่ i คือ

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}$$

โดยที่ $a(i)$ แทน ระยะห่างเฉลี่ยระหว่างข้อมูลที่ i กับข้อมูลอื่น ๆ ทั้งหมดภายในกลุ่มเดียวกัน

$b(i)$ แทน ระยะห่างเฉลี่ยที่น้อยที่สุดของระยะห่างเฉลี่ยระหว่างข้อมูลที่ i กับข้อมูลทั้งหมดที่อยู่กลุ่มอื่น ๆ

ดังนั้น $-1 \le S(i) \le 1$ ซึ่งถ้า $S(i)$ มีค่าใกล้ 1 แสดงว่าข้อมูลที่ i ถูกจัดให้อยู่ในกลุ่มที่เหมาะสมแล้ว ถ้า $S(i)$ มีค่าใกล้ -1 แสดงว่าข้อมูลที่ i ควรอยู่ในกลุ่มอื่น และถ้า $S(i)$ มีค่าใกล้ 0 แสดงว่าอาจอยู่กลุ่มใดกลุ่มหนึ่งก็ได้

ในการวิจัยนี้จะคำนวณค่า Average silhouette width เพื่อสังเกตความแตกต่างของค่า Average silhouette width ของการวิเคราะห์กลุ่มด้วยมาตรวัดระยะห่างแบบต่าง ๆ เท่านั้น แต่ไม่นำมาเปรียบเทียบว่าการวิเคราะห์กลุ่มด้วยมาตรวัดระยะห่างแบบใดมีประสิทธิภาพที่ดีกว่า เนื่องจากค่า Average silhouette width คำนวณจากระยะห่างระหว่างข้อมูล ซึ่งการวิจัยนี้มีการใช้ระยะห่างที่แตกต่างกัน และระยะห่างที่คำนวณได้ในแต่ละวิธีนั้นมีพิสัยที่แตกต่างกันอีกด้วย

บทที่ 3
วิธีดำเนินการวิจัย

การวิจัยครั้งนี้ มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพในการวิเคราะห์กลุ่มด้วยอัลกอริทึมจัดกลุ่มโดยรอบมีดอยด์ ที่ใช้ระยะห่างแบบต่าง ๆ สำหรับข้อมูลแบบผสม ที่ประกอบไปด้วยตัวแปรเชิงปริมาณ ตัวแปรนามบัญญัติ และตัวแปรอันดับ ผู้วิจัยพิจารณาค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของ ค่า Purity ค่า Rand statistic ค่า Jaccard coefficient รวมถึงค่า Average silhouette width ที่ได้จากการวิเคราะห์กลุ่มข้อมูลลักษณะต่าง ๆ โดยใช้โปรแกรม R เวอร์ชัน 3.2.3 ในการจำลองข้อมูลและวิเคราะห์กลุ่ม โดยได้กำหนดขอบเขตการวิจัย และดำเนินการวิจัย ดังนี้

3.1 ขอบเขตการวิจัย

การวิจัยครั้งนี้เป็นการศึกษาเปรียบเทียบเทคนิคการวิเคราะห์กลุ่มข้อมูลแบบผสม โดยใช้การวัดระยะห่างที่แตกต่างกัน ซึ่งมีขอบเขตการศึกษาดังต่อไปนี้

1. ตัวแปรที่ศึกษา

   1.1 ตัวแปรอิสระ (Independent variables; $\boldsymbol{X}$) จำนวน p ตัว นั่นคือ

$$\boldsymbol{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}$$

แบ่งเป็น

1.1.1 ตัวแปรนามบัญญัติ จำนวน p_{nom} ตัว
โดยตัวแปรที่ l มีจำนวนประเภทเท่ากับ q_l ประเภท

1.1.2 ตัวแปรอันดับ จำนวน p_{ord} ตัว
โดยตัวแปรที่ l มีจำนวนอันดับเท่ากับ q_l อันดับ

1.1.3 ตัวแปรเชิงปริมาณ จำนวน p_{quan} ตัว

ดังนั้น $p = p_{nom} + p_{ord} + p_{quan}$ โดยที่ทุกตัวแปรแปลงมาจากตัวแปรเชิงปริมาณ $\boldsymbol{Z}$ ที่มีการแจกแจงแบบปกติหลายตัวแปร (Multivariate Normal Distribution) จำนวน p ตัว นั่นคือ

$$\boldsymbol{Z} \sim MVN_p(\boldsymbol{\mu}, \Sigma)$$

โดยที่ $\boldsymbol{Z} = \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_p \end{bmatrix}$, $\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix}$ และ $\sum = \begin{bmatrix} \sigma_{11} & & \sigma_{p1} \\ & \ddots & \\ \sigma_{p1} & & \sigma_{pp} \end{bmatrix}$

จากนั้นแปลงตัวแปรเชิงปริมาณ $Z_1,...,Z_{p_{nom}}$ เป็นตัวแปรนามบัญญัติ และแปลงตัวแปรเชิงปริมาณ $Z_{p_{nom+1}},...,Z_{p_{nom}+p_{ord}}$ เป็นตัวแปรอันดับ โดยกำหนดค่าความน่าจะเป็นที่จะมีประเภทหรืออันดับต่าง ๆ สำหรับตัวแปรนามบัญญัติหรือตัวแปรอันดับที่ l คือ $P(X_l = 1), P(X_l = 2),..., P(X_l = q_l - 1)$ และ $P(X_l = q_l)$

1.2 ตัวแปรตาม (Dependent variables; Y) เป็นตัวแปรนามบัญญัติ จำนวน 1 ตัว และมีค่า $k = 1, 2,..., K$ ซึ่ง K คือจำนวนกลุ่มของข้อมูล เพื่อกำหนดว่าข้อมูลใดอยู่กลุ่มใด

2. กำหนดขนาดข้อมูลต่อกลุ่ม (n) เท่ากับ 20 และ 100 แทนกลุ่มข้อมูลขนาดเล็กและขนาดใหญ่ ตามลำดับ และกำหนดจำนวนกลุ่ม (K) เท่ากับ 3 และ 5 กลุ่ม ดังนั้นขนาดตัวอย่าง ($N = nK$) ที่เป็นไปได้ คือ 60 100 300 และ 500

3. กำหนดจำนวนตัวแปรอิสระ

3.1 จำนวนตัวแปรนามบัญญัติ p_{nom} = 0 และ 3

3.2 จำนวนตัวแปรอันดับ p_{ord} = 0 และ 3

3.3 จำนวนตัวแปรเชิงปริมาณ p_{quan} = 0 และ 3

และศึกษาข้อมูลที่มีชุดตัวแปร $p = p_{nom} + p_{ord} + p_{quan}$ ทั้งหมด 6 รูปแบบ ดังตารางที่ 3.1

ตารางที่ 3.1 ลักษณะชุดตัวแปรที่ศึกษารูปแบบต่าง ๆ

รูปแบบที่	p	p_{nom}	p_{ord}	p_{quan}
I	9	3	3	3
II	6	3	3	0
III	6	3	0	3
IV	6	0	3	3
V	3	3	0	0
VI	3	0	3	0

ซึ่งก็คือกลุ่มข้อมูลที่ประกอบไปด้วยตัวแปรอิสระชนิดต่าง ๆ 6 รูปแบบ ดังนี้

รูปแบบที่ I ข้อมูลที่ประกอบไปด้วย ตัวแปรนามบัญญัติ ตัวแปรอันดับ และตัวแปรเชิงปริมาณ อย่างละ 3 ตัวแปร

รูปแบบที่ II ข้อมูลที่ประกอบไปด้วย ตัวแปรนามบัญญัติ และตัวแปรอันดับ อย่างละ 3 ตัวแปร

รูปแบบที่ III ข้อมูลที่ประกอบไปด้วย ตัวแปรนามบัญญัติ และตัวแปรเชิงปริมาณ อย่างละ 3 ตัวแปร

รูปแบบที่ IV ข้อมูลที่ประกอบไปด้วย ตัวแปรอันดับ และตัวแปรเชิงปริมาณ อย่างละ 3 ตัวแปร

รูปแบบที่ V ข้อมูลที่ประกอบไปด้วย ตัวแปรนามบัญญัติ 3 ตัวแปร

รูปแบบที่ VI ข้อมูลที่ประกอบไปด้วย ตัวแปรอันดับ 3 ตัวแปร

4. กำหนดจำนวนประเภทหรืออันดับ (q_l) สำหรับตัวแปรนามบัญญัติหรือตัวแปรอันดับที่ l เท่ากับ 5 ดังนั้น ตัวแปรนามบัญญัติหรือตัวแปรอันดับ X_l แทนด้วย 1 2 3 4 และ 5 โดยความน่าจะเป็นดังตารางที่ 3.2 ซึ่งกำหนดให้ในกลุ่มที่ 1 มีความน่าจะเป็นที่ตัวแปร X_l เกิดขึ้นมากที่สุด หรือ $P(X_l=1)$ สูงสุด ขณะที่กลุ่มที่ 2 3 4 และ 5 จะมี $P(X_l=2)$ $P(X_l=3)$ $P(X_l=4)$ และ $P(X_l=5)$ สูงสุด ตามลำดับ เพื่อให้ข้อมูลตัวแปรนามบัญญัติหรือตัวแปรอันดับมีความแตกต่างระหว่างกลุ่ม

ตารางที่ 3.2 ความน่าจะเป็นของตัวแปรนามบัญญัติหรือตัวแปรอันดับ X_l ที่จะเกิดขึ้นในประเภทหรืออันดับต่าง ๆ 5 ประเภทหรืออันดับ

กลุ่มที่	$P(X_l=1)$	$P(X_l=2)$	$P(X_l=3)$	$P(X_l=4)$	$P(X_l=5)$	$\sum_{i=1}^{5}P(X_l=i)$
1	0.70	0.10	0.10	0.05	0.05	1.00
2	0.05	0.70	0.10	0.10	0.05	1.00
3	0.05	0.05	0.70	0.10	0.10	1.00
4	0.10	0.05	0.05	0.70	0.10	1.00
5	0.10	0.10	0.05	0.05	0.70	1.00

5. กำหนดเวกเตอร์ค่าเฉลี่ย

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_{p_{nom}} \\ \mu_{p_{nom}+1} \\ \vdots \\ \mu_{p_{nom}+p_{ord}} \\ \mu_{p_{nom}+p_{ord}+1} \\ \vdots \\ \mu_p \end{bmatrix}$$

เนื่องจากตัวแปรเชิงปริมาณ $Z_1,...,Z_{p_{nom}}$ จะถูกแปลงเป็นตัวแปรนามบัญญัติ และตัวแปรเชิงปริมาณ $Z_{p_{nom+1}},...,Z_{p_{nom}+p_{ord}}$ จะถูกแปลงเป็นตัวแปรอันดับ จึงกำหนดให้ค่าเฉลี่ย

$$\mu_1 = ... = \mu_{p_{nom}} = \mu_{p_{nom}+1} = ... = \mu_{p_{nom}+p_{ord}} = 0$$

กรณีที่จำนวนตัวแปรเชิงปริมาณ $p_{quan} = 3$ จึงได้ว่า

5.1 จำนวนกลุ่ม $K = 3$

$$\boldsymbol{\mu}_{31} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 2 \\ 3 \end{bmatrix}, \ \boldsymbol{\mu}_{32} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 4 \\ 5 \\ 6 \end{bmatrix}, \ \boldsymbol{\mu}_{33} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 7 \\ 8 \\ 9 \end{bmatrix}$$

5.2 จำนวนกลุ่ม $K = 5$

$$\boldsymbol{\mu}_{51} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 2 \\ 3 \end{bmatrix}, \ \boldsymbol{\mu}_{52} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 4 \\ 5 \\ 6 \end{bmatrix}, \ \boldsymbol{\mu}_{53} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 7 \\ 8 \\ 9 \end{bmatrix}, \ \boldsymbol{\mu}_{54} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 10 \\ 11 \\ 12 \end{bmatrix}, \ \boldsymbol{\mu}_{55} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 13 \\ 14 \\ 15 \end{bmatrix}$$

6. กำหนดความแปรปรวน

$$\sigma^2(Z_1) = \sigma^2(Z_2) = ... = \sigma^2(Z_p) = 1$$

7. กำหนดเมทริกซ์ของสัมประสิทธิ์สหสัมพันธ์ (Correlation matrix) ของ $\boldsymbol{Z}$ 2 กรณี ได้แก่

   7.1 ตัวแปรมีความสัมพันธ์ไปในทิศทางเดียวกัน น้อย

$$\boldsymbol{\rho} = \begin{bmatrix} 1.0 & 0.2 & \cdots & 0.2 \\ 0.2 & 1.0 & \cdots & 0.2 \\ \vdots & \vdots & \ddots & \vdots \\ 0.2 & 0.2 & \cdots & 1.0 \end{bmatrix}_{p \times p}$$

   7.2 ตัวแปรมีความสัมพันธ์ไปในทิศทางเดียวกัน มาก

$$\boldsymbol{\rho} = \begin{bmatrix} 1.0 & 0.8 & \cdots & 0.8 \\ 0.8 & 1.0 & \cdots & 0.8 \\ \vdots & \vdots & \ddots & \vdots \\ 0.8 & 0.8 & \cdots & 1.0 \end{bmatrix}_{p \times p}$$

8. กำหนดการจำลองข้อมูล 1,000 ครั้ง ต่อ 1 กรณี
9. มาตรวัดระยะห่างที่ใช้ในการศึกษาเป็นระยะห่างสำหรับข้อมูลแบบผสม ซึ่งมี 4 วิธีดังนี้

   9.1 ระยะห่างของ Kaufman and Rousseeuw (KR)

   9.2 ระยะห่างของ Podani (P)

   9.3 ระยะห่างแบบ Kaufman and Rousseeuw ร่วมกับ Noorbehbahani et al. (KR&N)

   9.4 ระยะห่างแบบ Podani ร่วมกับ Noorbehbahani et al. (P&N)

10. อัลกอริทึมที่ใช้ในการศึกษาการวิเคราะห์กลุ่มสำหรับข้อมูลแบบผสม คือ อัลกอริทึมจัดกลุ่มโดยรอบมีดอยด์ ด้วยฟังก์ชัน pam แพ็กเกจ cluster ในโปรแกรม R เวอร์ชัน 3.2.3
11. กำหนดข้อมูลจริง

    11.1 ข้อมูล **Pittsburgh Bridges Data Set**

    จากเว็บไซต์ https://archive.ics.uci.edu/ml/datasets/Pittsburgh+Bridges

    ซึ่งเป็นข้อมูลแบบผสม จำนวน 108 ชุดข้อมูล

    ประกอบไปด้วย
    - ตัวแปรตาม 1 ตัวแปร
    - ตัวแปรอิสระ 11 ตัวแปร

      ได้แก่

      ตัวแปรนามบัญญัติ 7 ตัวแปร

      ตัวแปรเชิงปริมาณ 3 ตัวแปร

      ตัวแปรอันดับ 1 ตัวแปร

12. กำหนดค่าวัดประสิทธิภาพการวิเคราะห์กลุ่ม ดังนี้

    12.1 ค่า Purity

    12.2 ค่า Rand statistic

    12.3 ค่า Jaccard coefficient

    12.4 ค่า Average silhouette width

13. กำหนดการทดสอบความแตกต่างของค่าเฉลี่ยของค่าวัดประสิทธิภาพการวิเคราะห์กลุ่ม ดังนี้

    กำหนดให้ μ_i = ค่าเฉลี่ยของค่า Purity ค่า Rand statistic ค่า Jaccard coefficient หรือ ค่า Average silhouette width ที่ได้จากการวิเคราะห์กลุ่มด้วยระยะห่าง i

    โดยที่ i = ระยะห่างของ KR ระยะห่างของ P ระยะห่างแบบ KR&N หรือ ระยะห่างแบบ P&N

    13.1 การวิเคราะห์ความแปรปรวนแบบมีปัจจัยเดียว (1-way ANOVA)

    หรือเรียกโดยย่อว่า การวิเคราะห์ความแปรปรวน เพื่อทดสอบความแตกต่างของค่าเฉลี่ยของค่าวัดประสิทธิภาพการวิเคราะห์กลุ่มข้อมูลรูปแบบที่ I และ II เนื่องจากเป็นข้อมูลที่มีการเปรียบประสิทธิภาพการวิเคราะห์กลุ่มด้วยระยะห่างทั้ง 4 วิธี โดยมีสมมติฐานเพื่อการทดสอบ ดังนี้

    สมมติฐาน $H_0 : \mu_{KR} = \mu_P = \mu_{KR\&N} = \mu_{P\&N}$

    H_1 : มีค่าเฉลี่ยอย่างน้อย 1 คู่ มีค่าไม่เท่ากัน

    กำหนดระดับนัยสำคัญ เท่ากับ 0.05

    13.2 การทดสอบความแตกต่างระหว่างค่าเฉลี่ยแต่ละคู่ของค่าที่วัดได้ ด้วยวิธี Tukey เพื่อหาประชากรที่มีค่าเฉลี่ยไม่เท่ากัน เมื่อเกิดการปฏิเสธสมมติฐานหลักในการวิเคราะห์ความแปรปรวน ซึ่งมีสมมติฐานเพื่อการทดสอบ ดังนี้

    สมมติฐาน $H_0 : \mu_i = \mu_j$

    $H_1 : \mu_i \neq \mu_j$

    กำหนดระดับนัยสำคัญ เท่ากับ 0.05

    13.3 การทดสอบความแตกต่างระหว่างค่าเฉลี่ยแต่ละคู่ของค่าที่วัดได้ ด้วยสถิติทดสอบ t เพื่อทดสอบความแตกต่างของค่าเฉลี่ยของค่าวัดประสิทธิภาพการวิเคราะห์กลุ่มข้อมูลรูปแบบที่ III IV V และ VI เนื่องจากเป็นข้อมูลที่มีการเปรียบประสิทธิภาพการวิเคราะห์กลุ่มด้วยระยะห่างเพียง 2 วิธี โดยมีสมมติฐานเพื่อการทดสอบ ดังนี้

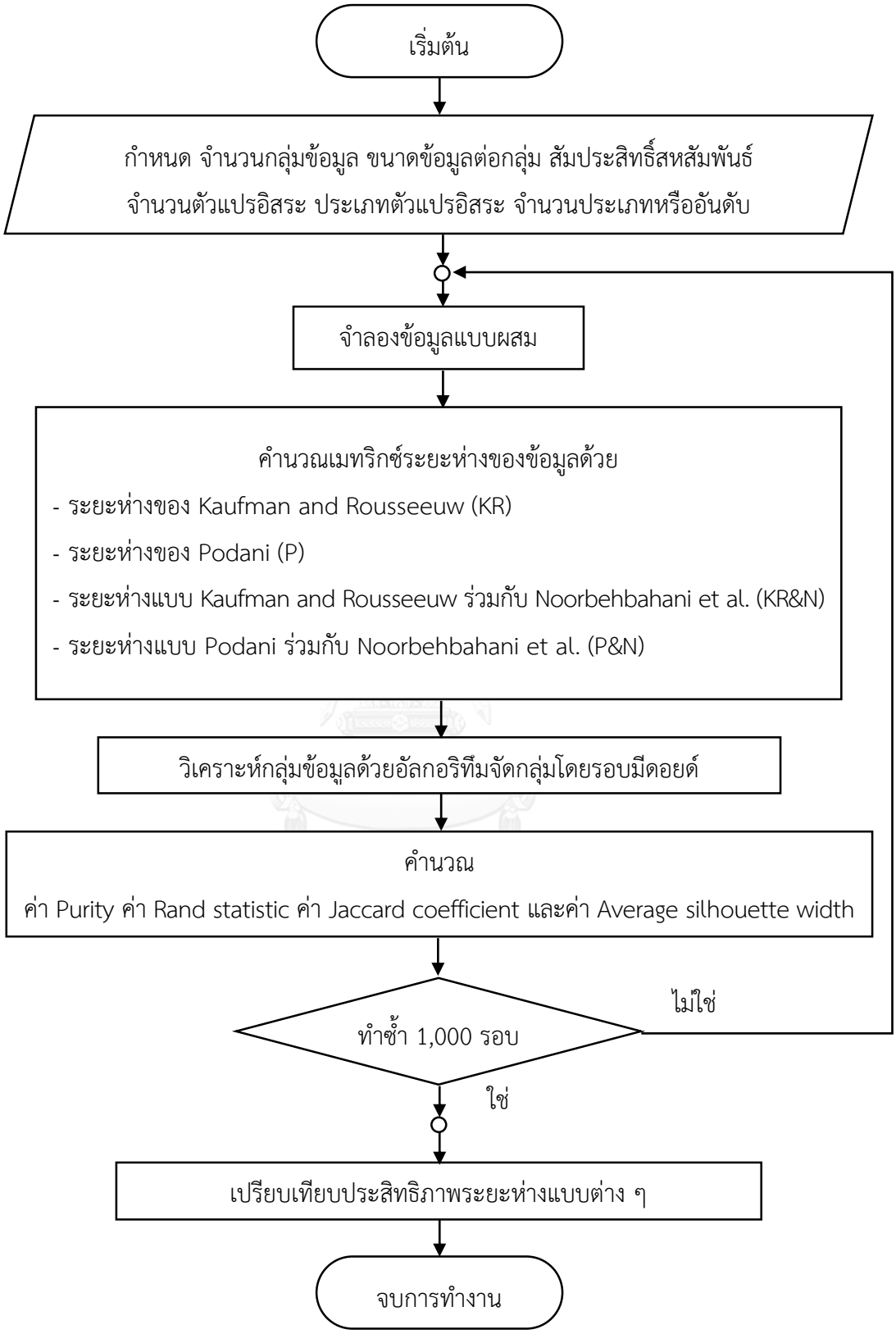
    สมมติฐาน $H_0 : \mu_i = \mu_j$

    $H_1 : \mu_i \neq \mu_j$

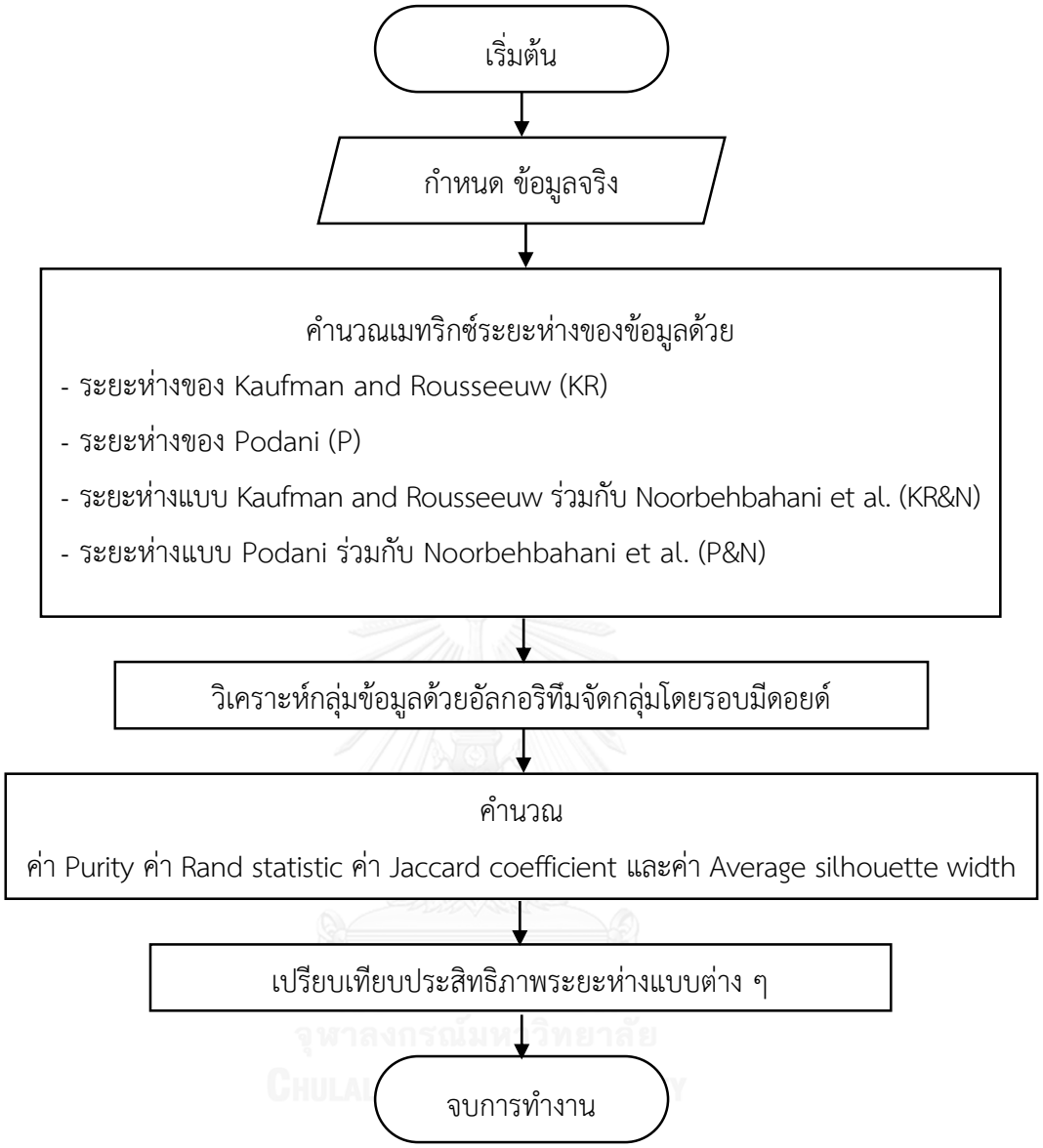
    กำหนดระดับนัยสำคัญ เท่ากับ 0.05

3.2 วิธีดำเนินการวิจัย

1. ศึกษามาตรวัดระยะห่างของข้อมูลแบบผสม
2. กำหนดค่าเริ่มต้นสำหรับจำลองข้อมูลแบบผสม ดังนี้
   2.1 จำนวนกลุ่มข้อมูล
   2.2 ขนาดข้อมูลต่อกลุ่ม
   2.3 ค่าสัมประสิทธิ์สหสัมพันธ์
   2.4 จำนวนตัวแปรอิสระ
   2.5 ชนิดตัวแปรอิสระ
   2.6 จำนวนประเภทหรืออันดับ
3. ทำการจำลองข้อมูลแบบผสมตามลักษณะที่กำหนดไว้ข้างต้น จำนวน 1,000 ครั้ง
4. นำข้อมูลที่จำลองขึ้นมาหาเมทริกซ์ระยะห่างด้วยมาตรวัดระยะห่าง ดังนี้
   4.1 ระยะห่างของ Kaufman and Rousseeuw (KR)
   4.2 ระยะห่างของ Podani (P)
   4.3 ระยะห่างแบบ Kaufman and Rousseeuw ร่วมกับ Noorbehbahani et al. (KR&N)
   4.4 ระยะห่างแบบ Podani ร่วมกับ Noorbehbahani et al. (P&N)
5. วิเคราะห์กลุ่มข้อมูลที่จำลองขึ้นด้วยอัลกอริทึมจัดกลุ่มโดยรอบมีดอยด์ โดยใช้เมทริกซ์ระยะห่างจากขั้นตอนที่ 4
6. คำนวณค่าวัดประสิทธิภาพการวิเคราะห์กลุ่มต่าง ๆ ดังนี้
   6.1 ค่า Purity
   6.2 ค่า Rand statistic
   6.3 ค่า Jaccard coefficient
   6.4 ค่า Average silhouette width
7. เปรียบเทียบประสิทธิภาพระยะห่างจากผลการวิเคราะห์กลุ่มในขั้นตอนที่ 6 โดยพิจารณา
   7.1 ค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐาน
   7.2 กราฟช่วงความเชื่อมั่น 95%
   7.3 การวิเคราะห์ความแปรปรวน
   7.4 การทดสอบความแตกต่างของค่าเฉลี่ยแต่ละคู่ของค่าที่วัดได้ ด้วยวิธี Tukey
   7.5 การทดสอบความแตกต่างของค่าเฉลี่ยแต่ละคู่ของค่าที่วัดได้ ด้วยสถิติทดสอบ t
8. กำหนดข้อมูลจริงที่จะใช้ในการศึกษา
9. วิเคราะห์กลุ่มข้อมูลจริงด้วยระยะห่างทั้ง 4 วิธี และศึกษาผลการวิเคราะห์กลุ่มข้อมูลจริง



รูปที่ 3.1 แผนผังขั้นตอนวิจัย สำหรับข้อมูลที่จำลองขึ้น



รูปที่ 3.2 แผนผังขั้นตอนวิจัย สำหรับข้อมูลจริง

บทที่ 4
ผลการวิเคราะห์ข้อมูล

ในการวิจัยครั้งนี้ ผู้วิจัยวิเคราะห์กลุ่มด้วยอัลกอริทึมจัดกลุ่มโดยรอบมีดอยด์ ซึ่งในที่นี้เรียกโดยย่อว่า “การวิเคราะห์กลุ่ม” โดยใช้มาตรวัดระยะห่างแบบต่าง ๆ 4 วิธี ได้แก่ ระยะห่างของ KR ระยะห่างของ P ระยะห่างแบบ KR&N และระยะห่างแบบ P&N เพื่อเปรียบเทียบประสิทธิภาพระยะห่างจากการวิเคราะห์กลุ่มข้อมูล ทั้งข้อมูลที่จำลองขึ้นและข้อมูลจริง โดยใช้ค่า Purity ค่า Rand statistic ค่า Jaccard coefficient และค่า Average silhouette width

4.1 ผลการวิเคราะห์กลุ่มข้อมูลที่จำลองขึ้น

ข้อมูลที่จำลองขึ้นมีลักษณะที่แตกต่างกันตามชนิดและจำนวนตัวแปรอิสระ ได้แก่ ตัวแปรนามบัญญัติ ตัวแปรอันดับ และตัวแปรเชิงปริมาณ โดยตัวแปรนามบัญญัติและตัวแปรอันดับ มีจำนวนประเภทหรืออันดับ เท่ากับ 5 ประเภทหรืออันดับ

ผู้วิจัยทำการวิเคราะห์กลุ่มข้อมูลที่ประกอบไปด้วยตัวแปรอิสระชนิดต่าง ๆ 6 รูปแบบ ดังนี้

รูปแบบที่ I ข้อมูลที่ประกอบไปด้วย ตัวแปรนามบัญญัติ ตัวแปรอันดับ และตัวแปรเชิงปริมาณ อย่างละ 3 ตัวแปร

รูปแบบที่ II ข้อมูลที่ประกอบไปด้วย ตัวแปรนามบัญญัติ และตัวแปรอันดับ อย่างละ 3 ตัวแปร

รูปแบบที่ III ข้อมูลที่ประกอบไปด้วย ตัวแปรนามบัญญัติ และตัวแปรเชิงปริมาณ อย่างละ 3 ตัวแปร

รูปแบบที่ IV ข้อมูลที่ประกอบไปด้วย ตัวแปรอันดับ และตัวแปรเชิงปริมาณ อย่างละ 3 ตัวแปร

รูปแบบที่ V ข้อมูลที่ประกอบไปด้วย ตัวแปรนามบัญญัติ 3 ตัวแปร

รูปแบบที่ VI ข้อมูลที่ประกอบไปด้วย ตัวแปรอันดับ 3 ตัวแปร

ซึ่งข้อมูลแต่ละรูปแบบถูกแบ่งอีกเป็นกรณีที่แตกต่างกันตามจำนวนกลุ่มข้อมูล (K = 3, 5) ค่าสัมประสิทธิ์สหสัมพันธ์ (ρ = 0.2, 0.8) และขนาดข้อมูลต่อกลุ่ม (n = 20, 100) จึงได้ข้อมูลแต่ละรูปแบบที่แตกต่างกันอีก 8 กรณี ดังนี้

กรณีที่ 1 เมื่อ K = 3, ρ = 0.2 และ n = 20

กรณีที่ 2 เมื่อ K = 3, ρ = 0.2 และ n = 100

กรณีที่ 3 เมื่อ K = 3, ρ = 0.8 และ n = 20

กรณีที่ 4 เมื่อ K = 3, ρ = 0.8 และ n = 100

กรณีที่ 5 เมื่อ K = 5, ρ = 0.2 และ n = 20

กรณีที่ 6 เมื่อ K = 5, ρ = 0.2 และ n = 100

กรณีที่ 7 เมื่อ K = 5, ρ = 0.8 และ n = 20

กรณีที่ 8 เมื่อ K = 5, ρ = 0.8 และ n = 100

ผู้วิจัยคำนวณค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของค่า Purity ค่า Rand statistic ค่า Jaccard coefficient และค่า Average silhouette width ที่ได้จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สร้างกราฟช่วงความเชื่อมั่น 95% นอกจากนี้ยังวิเคราะห์ความแปรปรวน และทดสอบความแตกต่างระหว่างค่าเฉลี่ยแต่ละคู่ของค่าที่วัดได้ด้วยวิธี Tukey เพื่อศึกษาความแตกต่างระหว่างประสิทธิภาพการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ ที่ระดับนัยสำคัญ 0.05 สำหรับข้อมูลรูปแบบที่ I และ II ซึ่งเป็นข้อมูลที่มีการวิเคราะห์กลุ่มด้วยระยะห่างทั้ง 4 วิธี เนื่องจากหากพิจารณาเพียงกราฟช่วงความเชื่อมั่น 95% เพื่อตัดสินความแตกต่างระหว่างประสิทธิภาพการวิเคราะห์กลุ่มด้วยระยะห่างทั้ง 4 วิธีนี้ อาจทำให้เกิดความผิดพลาดประเภทที่ 1 (Type 1 error) มาก เพราะว่าต้องเปรียบเทียบประสิทธิภาพการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ ถึง 6 คู่ ซึ่งเมื่อจำนวนคู่ที่จะเปรียบเทียบเพิ่มขึ้น ระดับนัยสำคัญจะเพิ่มขึ้นตาม เกิดช่วงความเชื่อมั่น 95% ที่แคบกว่าที่ควรจะเป็น ขณะที่การทดสอบความแตกต่างระหว่างค่าเฉลี่ยแต่ละคู่ของค่าที่วัดได้ด้วยวิธี Tukey ซึ่งเป็นการเปรียบเทียบเชิงพหุ (Multiple comparison) วิธีหนึ่ง โดยใช้ Studentized range distribution ทำให้ได้ช่วงความเชื่อมั่นที่กว้างขึ้น โอกาสที่จะปฏิเสธสมมติฐานหลักจึงลดลง

นอกจากนี้ ทำการทดสอบความแตกต่างระหว่างค่าเฉลี่ยแต่ละคู่ของค่าที่วัดได้ด้วยสถิติทดสอบ t สำหรับข้อมูลรูปแบบที่ III IV V และ VI เพื่อศึกษาความแตกต่างระหว่างประสิทธิภาพการวิเคราะห์กลุ่มที่ระดับนัยสำคัญ 0.05 เนื่องจากเป็นข้อมูลที่มีการเปรียบเทียบประสิทธิภาพการวิเคราะห์กลุ่มด้วยระยะห่างเพียง 2 วิธี หรือเพียง 1 คู่เท่านั้น

4.1.1 ผลการวิเคราะห์กลุ่มข้อมูลรูปแบบที่ I

ผู้วิจัยทำการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ 4 วิธี ได้แก่ ระยะห่างของ KR ระยะห่างของ P ระยะห่างแบบ KR&N และระยะห่างแบบ P&N สำหรับข้อมูลรูปแบบที่ I ซึ่งเป็นข้อมูลที่ประกอบไปด้วยตัวแปรนามบัญญัติ ตัวแปรอันดับ และตัวแปรเชิงปริมาณ อย่างละ 3 ตัวแปร ซึ่งแบ่งได้อีกเป็นกรณีที่แตกต่างกันตามจำนวนกลุ่มข้อมูล ค่าสัมประสิทธิ์สหสัมพันธ์ และขนาดข้อมูลต่อกลุ่ม คำนวณค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของค่า Purity ค่า Rand statistic ค่า Jaccard coefficient และค่า Average silhouette width ได้ผลดังตารางที่ 4.1

ผลการเปรียบเทียบประสิทธิภาพโดยเฉลี่ยในการวิเคราะห์กลุ่มด้วยมาตรวัดระยะห่างแบบต่าง ๆ จากตารางที่ 4.1 พบว่า

เมื่อ K = 3 การวิเคราะห์กลุ่มด้วยระยะห่างแบบ KR&N ให้ค่าเฉลี่ยของค่า Purity ค่า Rand statistic ค่า Jaccard coefficient สูงที่สุด ทุกกรณี แสดงว่าการวิเคราะห์กลุ่มด้วยระยะห่างแบบ KR&N มีประสิทธิภาพดีที่สุดโดยเฉลี่ย

เมื่อ K = 5 กรณีที่ ρ = 0.2 และ n = 20 การวิเคราะห์กลุ่มด้วยระยะห่างแบบ P&N ให้ค่าเฉลี่ยของค่า Purity ค่า Rand statistic ค่า Jaccard coefficient สูงที่สุด แสดงว่าการวิเคราะห์กลุ่มด้วยระยะห่างแบบ P&N มีประสิทธิภาพดีที่สุดโดยเฉลี่ย

เมื่อ K = 5 กรณีที่ ρ = 0.2 และ n = 100 และกรณีที่ ρ = 0.8, n = 20 และ 100 การวิเคราะห์กลุ่มด้วยระยะห่างแบบ KR&N และระยะห่างแบบ P&N ให้ค่าเฉลี่ยของค่า Purity ค่า Rand statistic และค่า Jaccard coefficient สูงใกล้เคียงกัน แสดงว่าการวิเคราะห์กลุ่มด้วยระยะห่างแบบ KR&N และ P&N มีประสิทธิภาพดีที่สุดโดยเฉลี่ยใกล้เคียงกัน

นอกจากนี้ ค่าเฉลี่ยของค่า Average silhouette width จากการวิเคราะห์กลุ่มด้วยระยะห่างของ P มีค่าสูงที่สุด ตามด้วยระยะห่างของ KR ระยะห่างแบบ P&N และระยะห่างแบบ KR&N ตามลำดับ ในทุกกรณี ซึ่งแตกต่างจากค่าวัดประสิทธิภาพการวิเคราะห์กลุ่มอื่น ๆ

ตารางที่ 4.1 ค่าเฉลี่ย (ส่วนเบี่ยงเบนมาตรฐาน) ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ I

K	ρ	n	ค่าที่วัด	ระยะห่าง			
				KR	P	KR&N	P&N
3	0.2	20	ค่า Purity	0.9526 (0.0281)	0.9531 (0.0282)	**0.9624** (0.0262)	0.9558 (0.0274)
			ค่า Rand statistic	0.9399 (0.0342)	0.9405 (0.0344)	**0.9523** (0.0320)	0.9439 (0.0334)
			ค่า Jaccard coefficient	0.8343 (0.0866)	0.8358 (0.0874)	**0.8665** (0.0836)	0.8445 (0.0855)
			ค่า Average silhouette width	0.4398 (0.0406)	**0.4460** (0.0408)	0.3818 (0.0346)	0.3934 (0.0366)
		100	ค่า Purity	0.9531 (0.0121)	0.9532 (0.0124)	**0.9663** (0.0113)	0.9570 (0.0122)
			ค่า Rand statistic	0.9400 (0.0149)	0.9400 (0.0153)	**0.9566** (0.0141)	0.9449 (0.0151)
			ค่า Jaccard coefficient	0.8349 (0.0377)	0.8350 (0.0387)	**0.8778** (0.0373)	0.8475 (0.0386)
			ค่า Average silhouette width	0.4383 (0.0186)	**0.4443** (0.0185)	0.3675 (0.0143)	0.3796 (0.0152)
	0.8	20	ค่า Purity	0.8344 (0.0464)	0.8319 (0.0471)	**0.8396** (0.0469)	0.8334 (0.0471)
			ค่า Rand statistic	0.8087 (0.0462)	0.8040 (0.0479)	**0.8165** (0.0465)	0.8074 (0.0473)
			ค่า Jaccard coefficient	0.5534 (0.0830)	0.5464 (0.0840)	**0.5681** (0.0841)	0.5524 (0.0834)
			ค่า Average silhouette width	0.5143 (0.0570)	**0.5287** (0.0567)	0.4473 (0.0504)	0.4704 (0.0529)
		100	ค่า Purity	0.8326 (0.0228)	0.8286 (0.0227)	**0.8383** (0.0233)	0.8313 (0.0229)
			ค่า Rand statistic	0.8057 (0.0230)	0.7991 (0.0231)	**0.8138** (0.0230)	0.8031 (0.0232)
			ค่า Jaccard coefficient	0.5518 (0.0401)	0.5415 (0.0394)	**0.5660** (0.0411)	0.5483 (0.0399)
			ค่า Average silhouette width	0.5160 (0.0256)	**0.5327** (0.0254)	0.4304 (0.0222)	0.4584 (0.0234)

ตารางที่ 4.1 (ต่อ) ค่าเฉลี่ย (ส่วนเบี่ยงเบนมาตรฐาน) ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ I

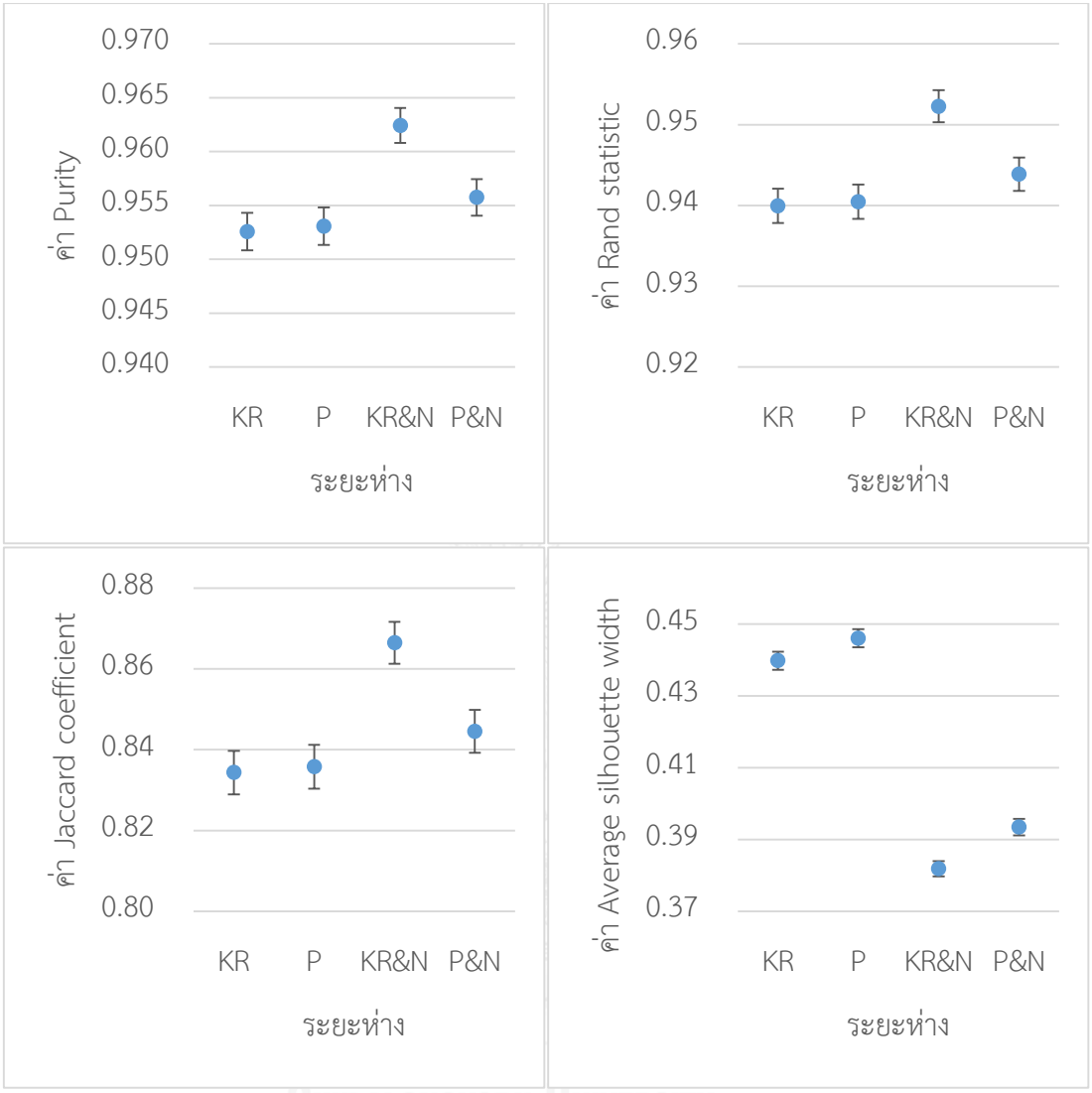
K	ρ	n	ค่าที่วัด	ระยะห่าง			
				KR	P	KR&N	P&N
5	0.2	20	ค่า Purity	0.9241 (0.0274)	0.9244 (0.0272)	0.9244 (0.0267)	**0.9246** (0.0267)
			ค่า Rand statistic	0.9433 (0.0193)	0.9436 (0.0192)	0.9435 (0.0189)	**0.9437** (0.0188)
			ค่า Jaccard coefficient	0.7471 (0.0755)	0.7479 (0.0752)	0.7478 (0.0735)	**0.7483** (0.0735)
			ค่า Average silhouette width	0.4085 (0.0344)	**0.4088** (0.0344)	0.4049 (0.0343)	0.4051 (0.0343)
		100	ค่า Purity	0.9245 (0.0116)	0.9246 (0.0116)	**0.9249** (0.0117)	**0.9249** (0.0116)
			ค่า Rand statistic	0.9428 (0.0083)	0.9429 (0.0083)	**0.9431** (0.0084)	**0.9431** (0.0083)
			ค่า Jaccard coefficient	0.7490 (0.0318)	0.7492 (0.0319)	**0.7501** (0.0321)	**0.7501** (0.0319)
			ค่า Average silhouette width	0.4089 (0.0147)	**0.4090** (0.0148)	0.4075 (0.0147)	0.4075 (0.0147)
	0.8	20	ค่า Purity	0.7749 (0.0426)	0.7750 (0.0427)	**0.7755** (0.0423)	**0.7755** (0.0424)
			ค่า Rand statistic	0.8487 (0.0238)	**0.8488** (0.0237)	**0.8488** (0.0237)	**0.8488** (0.0237)
			ค่า Jaccard coefficient	0.4417 (0.0629)	0.4418 (0.0628)	0.4424 (0.0626)	**0.4425** (0.0627)
			ค่า Average silhouette width	**0.5268** (0.0439)	**0.5268** (0.0439)	0.5199 (0.0445)	0.5203 (0.0445)
		100	ค่า Purity	0.7739 (0.0181)	0.7740 (0.0181)	0.7747 (0.0179)	**0.7748** (0.0179)
			ค่า Rand statistic	0.8460 (0.0102)	0.8461 (0.0103)	**0.8465** (0.0102)	**0.8465** (0.0101)
			ค่า Jaccard coefficient	0.4423 (0.0267)	0.4425 (0.0268)	0.4436 (0.0266)	**0.4437** (0.0265)
			ค่า Average silhouette width	**0.5314** (0.0189)	**0.5314** (0.0189)	0.5287 (0.0190)	0.5288 (0.0191)

พิจารณากราฟช่วงความเชื่อมั่น 95% ของค่าต่าง ๆ พบว่า

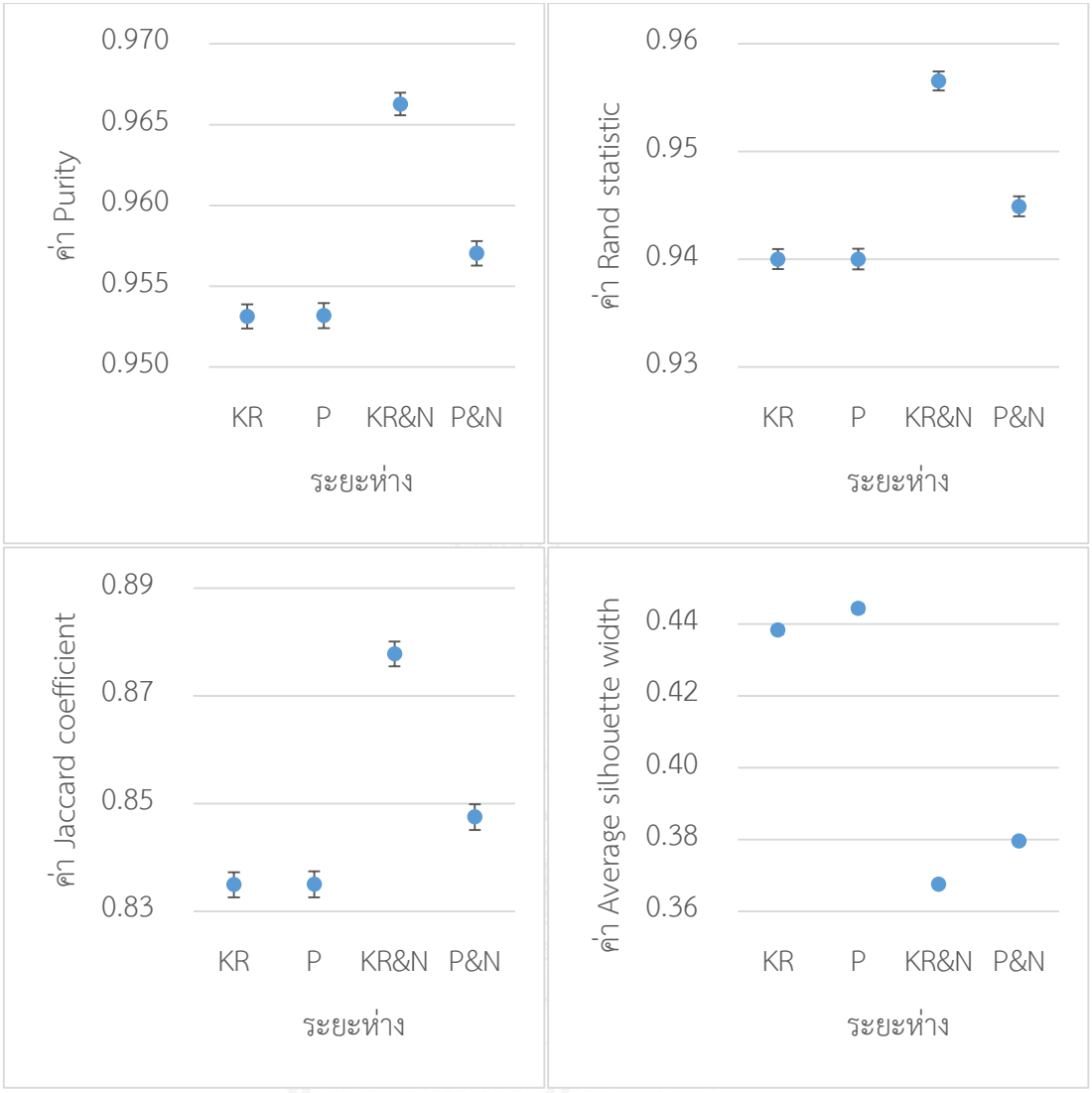
เมื่อ K = 3 ค่าเฉลี่ยของค่า Purity ค่า Rand statistic และค่า Jaccard coefficient จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบ KR&N มีค่ามากที่สุด โดยไม่มีส่วนของช่วงความเชื่อมั่น 95% ทับซ้อนกับค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างอีก 3 วิธี แสดงว่าการวิเคราะห์กลุ่มด้วยระยะห่างแบบ KR&N มีประสิทธิภาพดีที่สุดโดยเฉลี่ย และแตกต่างจากการวิเคราะห์กลุ่มด้วยระยะห่างแบบอื่น ๆ ซึ่งมีประสิทธิภาพรองลงมา ดังรูปที่ 4.1 ถึง 4.4 นอกจากนี้ เมื่อค่าสัมประสิทธิ์สหสัมพันธ์เพิ่มขึ้นจาก ρ = 0.2 เป็น ρ = 0.8 ระยะห่างที่ทำให้การวิเคราะห์กลุ่มมีประสิทธิภาพโดยเฉลี่ยในอันดับที่ 2 เปลี่ยนจากระยะห่างแบบ P&N เป็นระยะห่างของ KR โดยที่ค่าเฉลี่ยของค่า Purity ค่า Rand statistic และค่า Jaccard coefficient ของระยะห่างทั้งสองนี้มีส่วนของช่วงความเชื่อมั่น 95% ทับซ้อนกันมากขึ้น

เมื่อ K = 3 ค่า Average silhouette width จากการวิเคราะห์กลุ่มด้วยระยะห่างของ P ระยะห่างของ KR ระยะห่างแบบ P&N และระยะห่างแบบ KR&N มีค่าจากมากไปน้อยตามลำดับ และไม่มีส่วนของช่วงความเชื่อมั่น 95% ที่ทับซ้อนกันในทุกกรณี

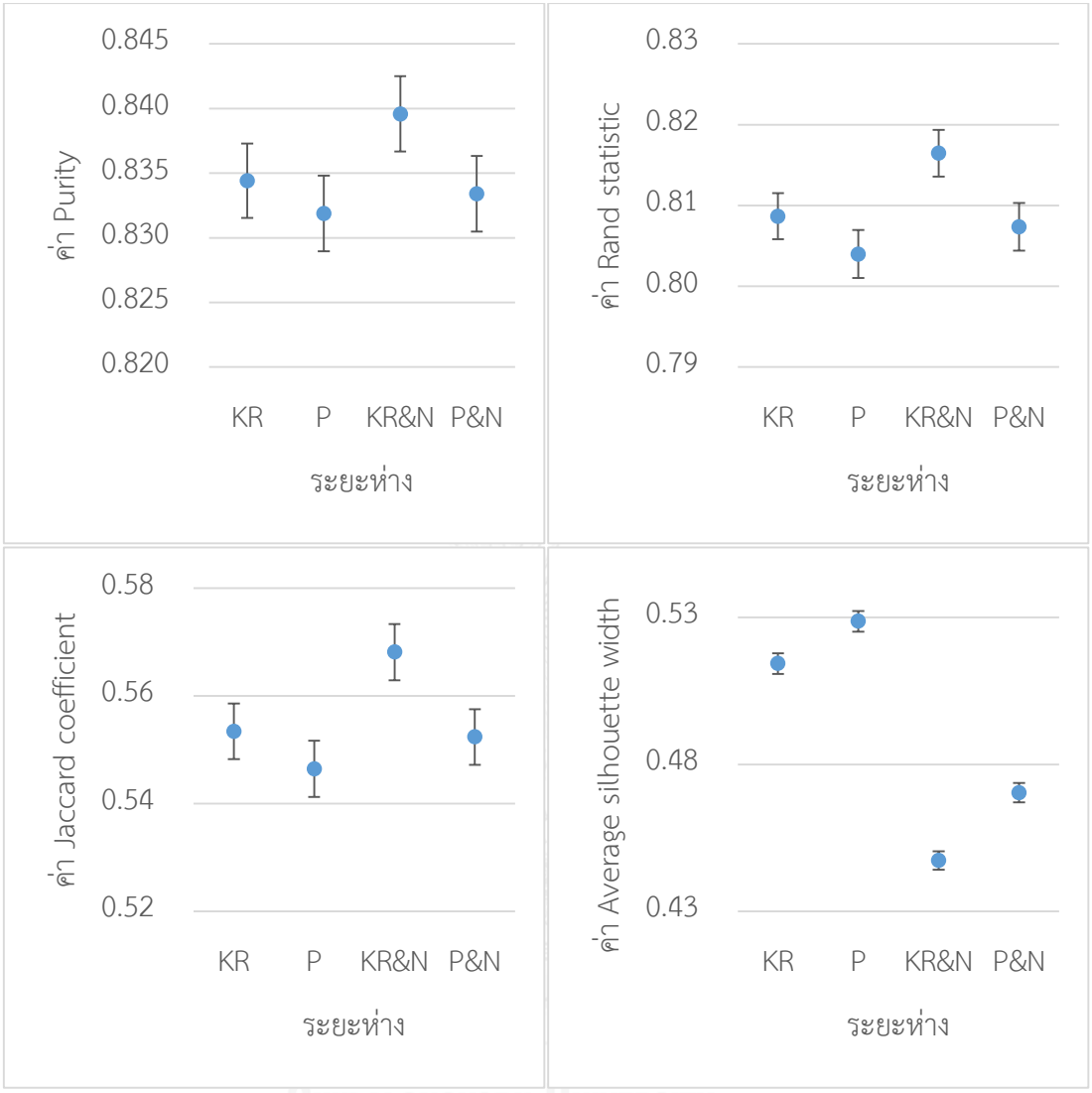
เมื่อ K = 5 แม้ว่าค่าเฉลี่ยของ ค่า Purity ค่า Rand statistic และค่า Jaccard coefficient จากการวิเคราะห์กลุ่มด้วยระยะห่างทั้ง 4 วิธี จะมีค่าแตกต่างกัน แต่เมื่อพิจารณากราฟช่วงความเชื่อมั่น 95% พบว่า การวิเคราะห์กลุ่มด้วยระยะห่างทั้ง 4 วิธี มีประสิทธิภาพโดยเฉลี่ยใกล้เคียงกันและมีส่วนของช่วงความเชื่อมั่น 95% ทับซ้อนกันในทุกกรณี นอกจากนี้ค่าเฉลี่ยของค่า Average silhouette width จากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR และระยะห่างของ P มีค่าใกล้เคียงกันและมากกว่า จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบ KR&N และระยะห่างแบบ P&N ซึ่งมีค่าใกล้เคียงกันเช่นกัน ดังรูปที่ 4.5 ถึง 4.8



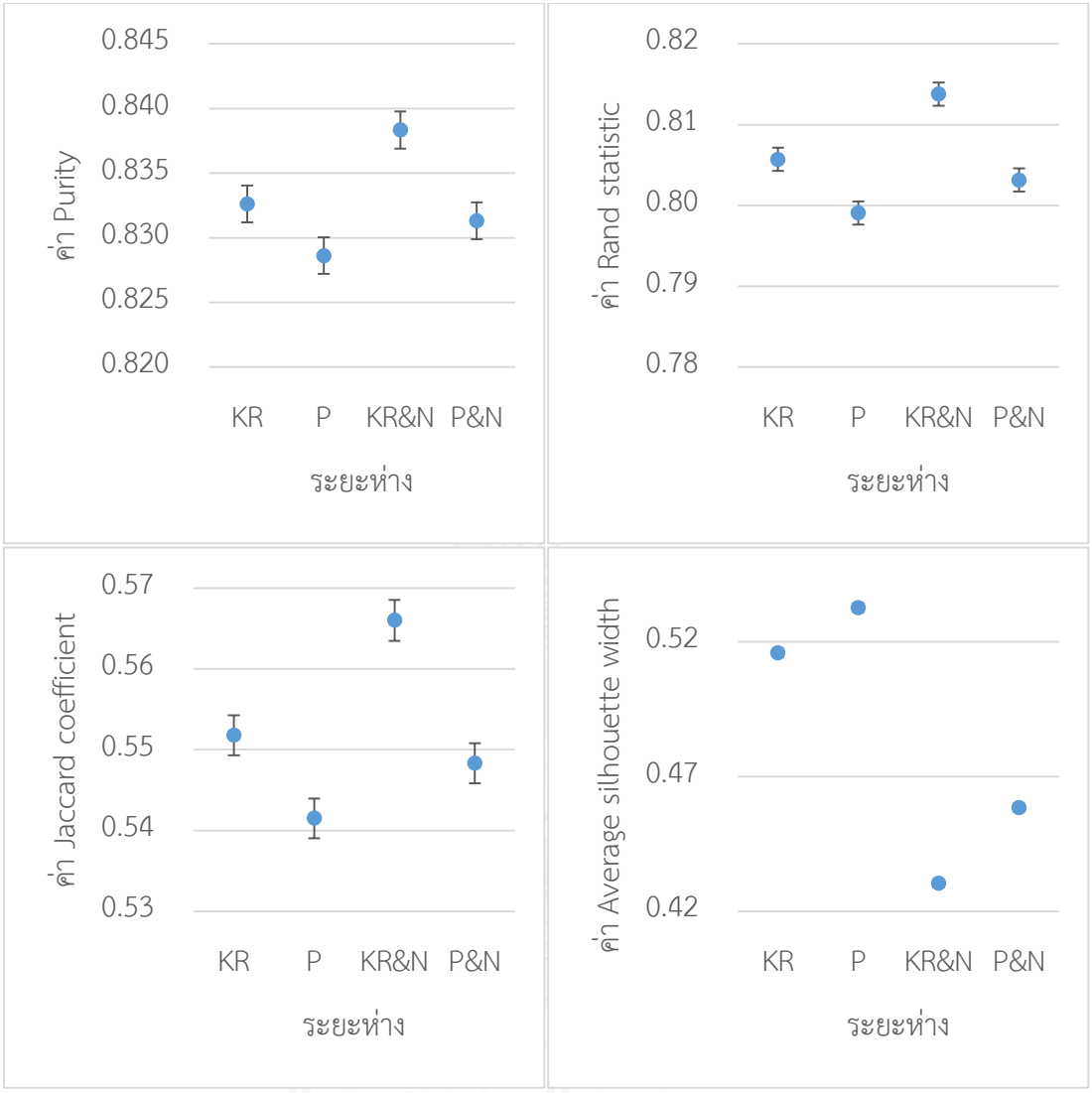
รูปที่ 4.1 กราฟช่วงความเชื่อมั่น 95% ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ I เมื่อ $K = 3$, $\rho = 0.2$ และ $n = 20$



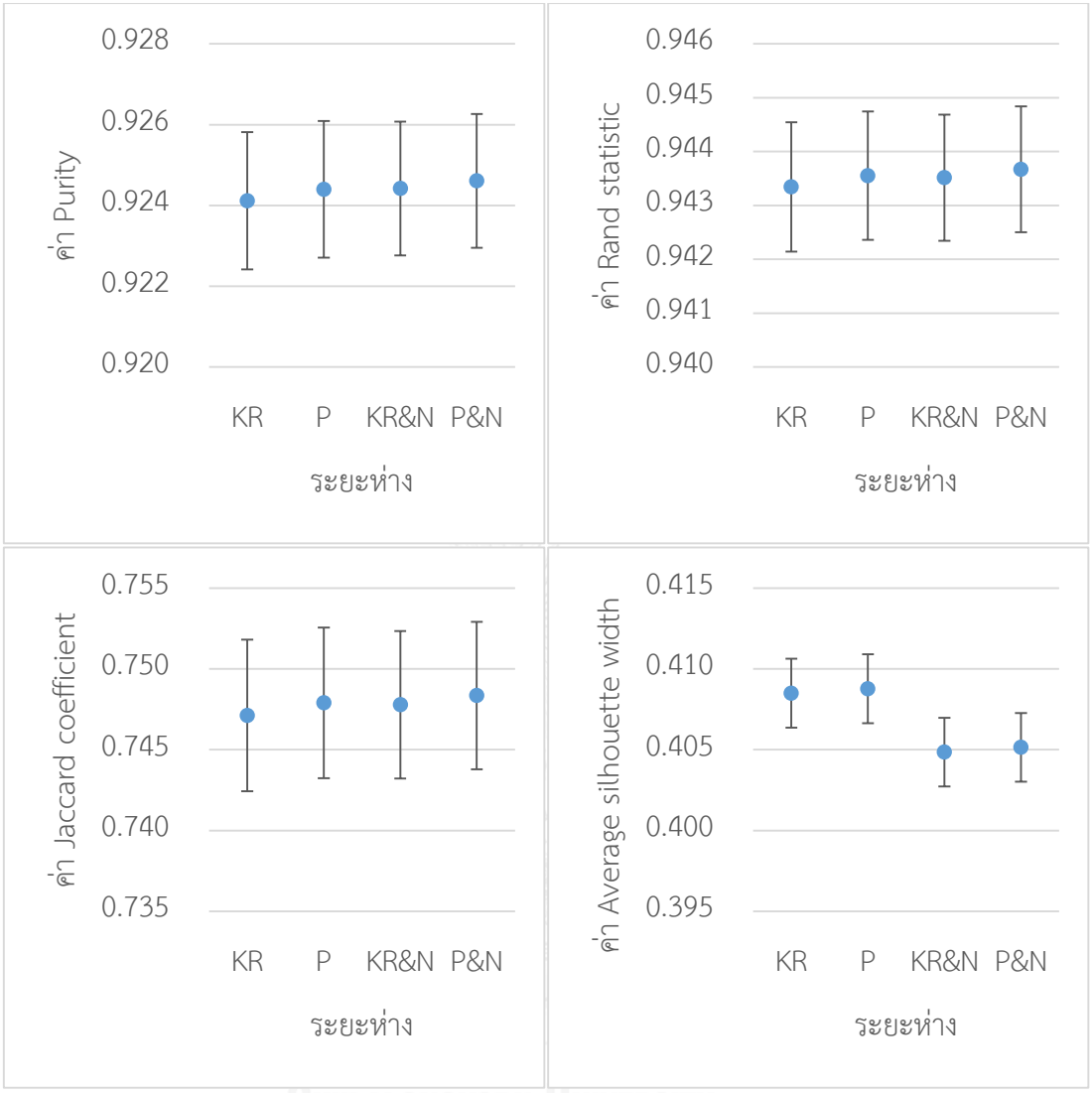
รูปที่ 4.2 กราฟช่วงความเชื่อมั่น 95% ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ I เมื่อ $K = 3$, $\rho = 0.2$ และ $n = 100$



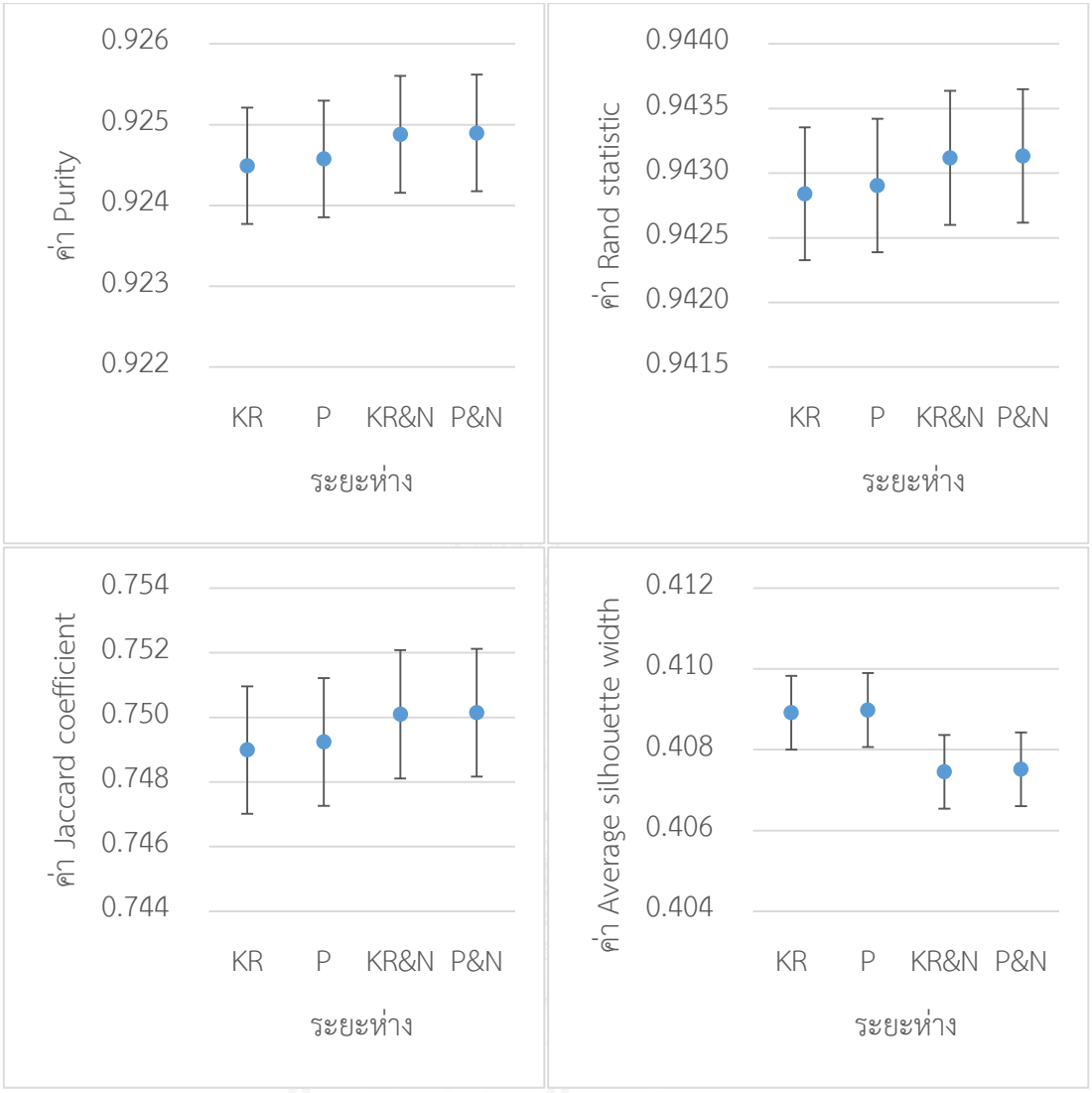
รูปที่ 4.3 กราฟช่วงความเชื่อมั่น 95% ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ I เมื่อ $K = 3$, $\rho = 0.8$ และ $n = 20$



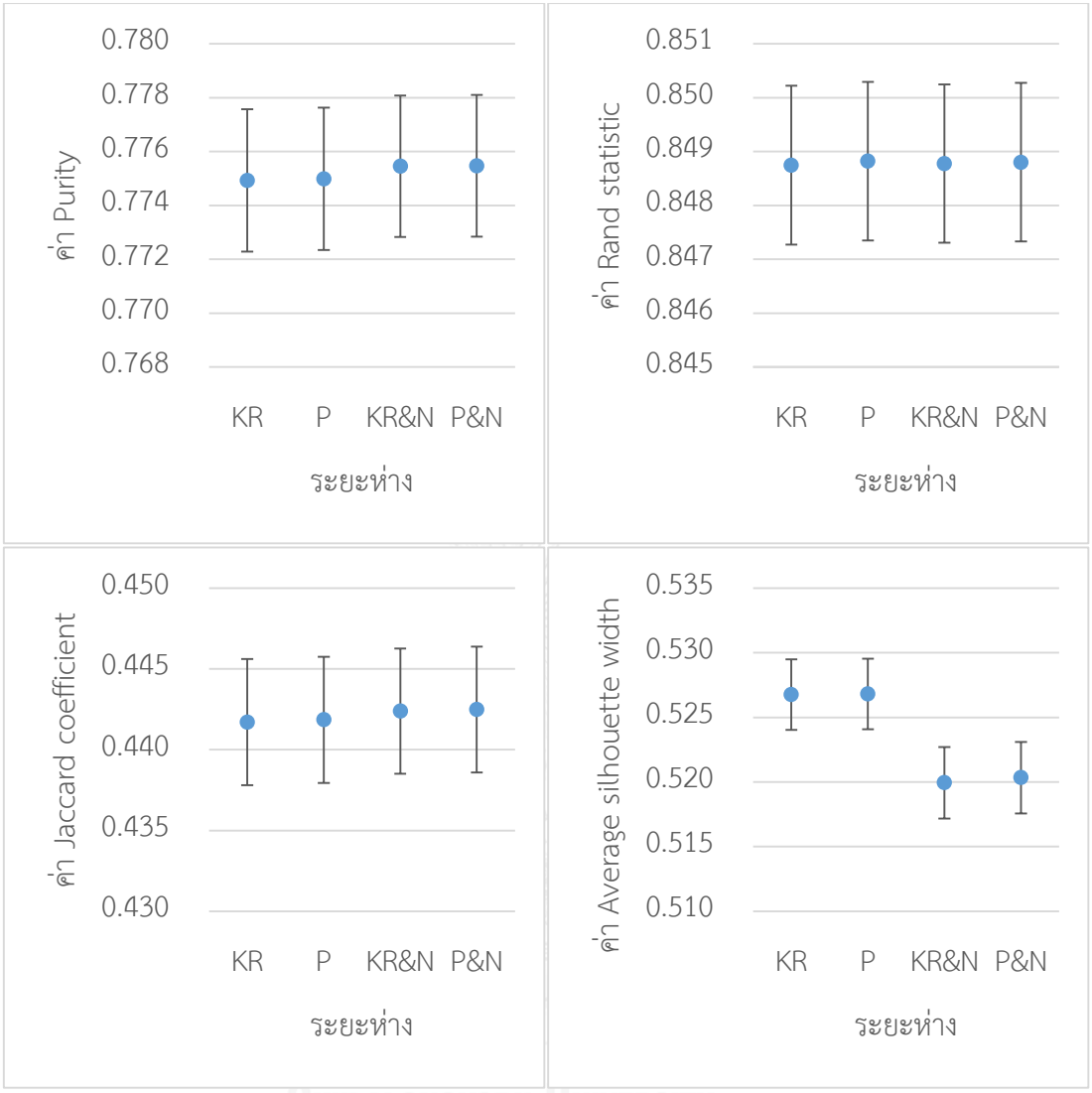
รูปที่ 4.4 กราฟช่วงความเชื่อมั่น 95% ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ I เมื่อ $K = 3$, $\rho = 0.8$ และ $n = 100$



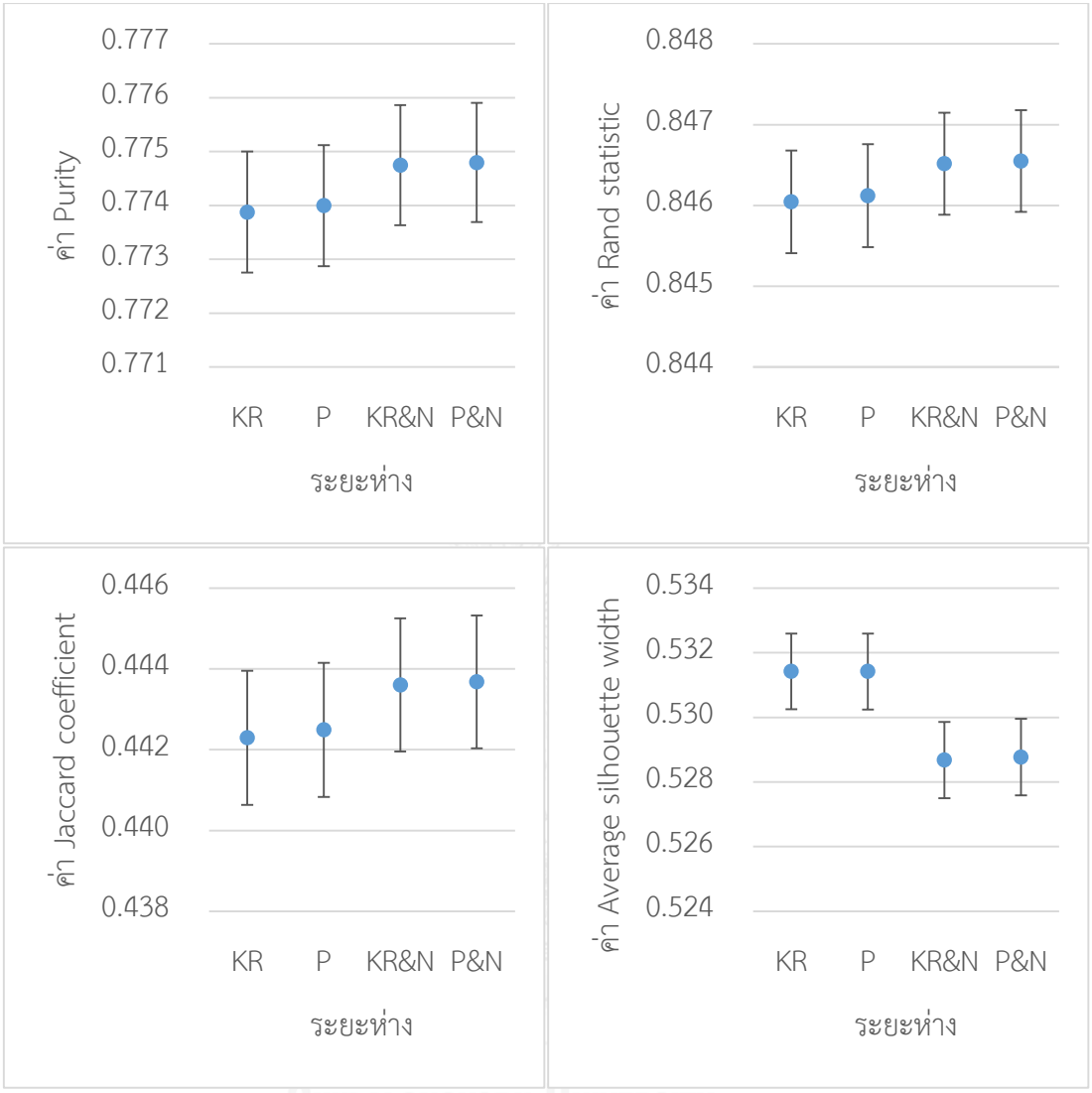
รูปที่ 4.5 กราฟช่วงความเชื่อมั่น 95% ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ I เมื่อ $K = 5$, $\rho = 0.2$ และ $n = 20$



รูปที่ 4.6 กราฟช่วงความเชื่อมั่น 95% ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ I เมื่อ $K = 5$, $\rho = 0.2$ และ $n = 100$



รูปที่ 4.7 กราฟช่วงความเชื่อมั่น 95% ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ I เมื่อ $K = 5$, $\rho = 0.8$ และ $n = 20$



รูปที่ 4.8 กราฟช่วงความเชื่อมั่น 95% ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ I เมื่อ $K = 5$, $\rho = 0.8$ และ $n = 100$

เนื่องจากการพิจารณากราฟช่วงความเชื่อมั่น 95% ของข้อมูลรูปแบบที่ I พบว่ามีบางกรณีที่ค่าเฉลี่ยของค่าวัดประสิทธิภาพการวิเคราะห์กลุ่มใกล้เคียงกัน และมีช่วงความเชื่อมั่น 95% ทับซ้อนกัน ผู้วิจัยจึงสนใจพิจารณาความแตกต่างของประสิทธิภาพการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ ที่ระดับนัยสำคัญ โดยทำการศึกษาข้อมูลรูปแบบที่ I แบ่งเป็นกรณีใหญ่ 2 กรณีตามจำนวนกลุ่มข้อมูล คือกรณีที่ K = 3 และ K = 5

4.1.1.1 ผลทดสอบความแตกต่างการวิเคราะห์กลุ่มข้อมูลรูปแบบที่ I เมื่อ K = 3

นอกจากข้อมูลจะแตกต่างกันตามจำนวนกลุ่มแล้ว ข้อมูลยังแตกต่างกันตามค่าสัมประสิทธิ์สหสัมพันธ์ และขนาดข้อมูลต่อกลุ่มอีกด้วย จึงทำการศึกษาประสิทธิภาพการวิเคราะห์กลุ่มข้อมูลต่าง ๆ ได้ดังต่อไปนี้

ข้อมูลรูปแบบที่ I เมื่อ K = 3, ρ = 0.2 และ n = 20

สำหรับข้อมูลรูปแบบที่ I เมื่อ K = 3, ρ = 0.2 และ n = 20 ทำการวิเคราะห์ความแปรปรวน พบว่าค่า Purity ค่า Rand statistic ค่า Jaccard coefficient และค่า Average silhouette width ต่างมี Sig. = 0.0000 ดังตารางที่ 4.2 ซึ่งน้อยกว่า 0.05 จึงปฏิเสธสมมติฐานหลัก นั่นคือ ประสิทธิภาพการวิเคราะห์กลุ่มขึ้นกับระยะห่างที่ใช้ในการวิเคราะห์กลุ่มที่ระดับนัยสำคัญ 0.05 และค่า Average silhouette width มีค่าขึ้นกับระยะห่างที่ใช้ในการวิเคราะห์กลุ่มที่ระดับนัยสำคัญ 0.05 เช่นเดียวกัน

ตารางที่ 4.2 ผลวิเคราะห์ความแปรปรวนของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ I เมื่อ $K = 3$, $\rho = 0.2$ และ $n = 20$

แหล่งความแปรปรวน		SS	df	MS	F	Sig.
ค่า Purity	ระหว่างกลุ่ม	0.0616	3	0.0205	27.1474	**0.0000**
	ภายในกลุ่ม	3.0237	3996	0.0008		
	ผลรวม	3.0853	3999			
ค่า Rand statistic	ระหว่างกลุ่ม	0.0974	3	0.0325	28.9644	**0.0000**
	ภายในกลุ่ม	4.4795	3996	0.0011		
	ผลรวม	4.5769	3999			
ค่า Jaccard coefficient	ระหว่างกลุ่ม	0.6594	3	0.2198	29.8633	**0.0000**
	ภายในกลุ่ม	29.4120	3996	0.0074		
	ผลรวม	30.0715	3999			
ค่า Average silhouette width	ระหว่างกลุ่ม	3.1413	3	1.0471	716.0752	**0.0000**
	ภายในกลุ่ม	5.8432	3996	0.0015		
	ผลรวม	8.9845	3999			

ผู้วิจัยจึงทดสอบความแตกต่างของค่าเฉลี่ยของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแต่ละคู่ พบว่าค่าเฉลี่ยของค่า Purity ค่า Rand statistic และค่า Jaccard coefficient ที่ได้จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบ KR&N ต่างก็แตกต่างจากระยะห่างวิธีอื่น ๆ ที่ระดับนัยสำคัญ 0.05 นั่นคือการวิเคราะห์กลุ่มด้วยระยะห่างแบบ KR&N มีประสิทธิภาพแตกต่างจากการวิเคราะห์กลุ่มด้วยระยะห่างอีก 3 วิธีอย่างมีนัยสำคัญ และยังพบว่าการวิเคราะห์กลุ่มด้วยระยะห่างอีก 3 วิธีนี้มีประสิทธิภาพไม่แตกต่างกัน นอกจากนี้ค่าเฉลี่ยของค่า Average silhouette width จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ ทั้ง 4 วิธีนี้ แตกต่างกันที่ระดับนัยสำคัญ 0.05 ดังตารางที่ 4.3

ตารางที่ 4.3 ผลทดสอบความแตกต่างระหว่างค่าเฉลี่ยแต่ละคู่ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ I เมื่อ K = 3, ρ = 0.2 และ n = 20 ด้วยวิธี Tukey

ค่าที่วัด	ระยะห่าง		ผลต่างค่าเฉลี่ย	Std. Error	Sig.
ค่า Purity	KR	P	-0.0005	0.0012	0.9773
		KR&N	-0.0099	0.0012	**0.0000**
		P&N	-0.0032	0.0012	0.0495
	P	KR	0.0005	0.0012	0.9773
		KR&N	-0.0094	0.0012	**0.0000**
		P&N	-0.0027	0.0012	0.1325
	KR&N	KR	0.0099	0.0012	**0.0000**
		P	0.0094	0.0012	**0.0000**
		P&N	0.0067	0.0012	**0.0000**
	P&N	KR	0.0032	0.0012	0.0495
		P	0.0027	0.0012	0.1325
		KR&N	-0.0067	0.0012	**0.0000**
ค่า Rand statistic	KR	P	-0.0005	0.0015	0.9857
		KR&N	-0.0123	0.0015	**0.0000**
		P&N	-0.0039	0.0015	0.0433
	P	KR	0.0005	0.0015	0.9857
		KR&N	-0.0118	0.0015	**0.0000**
		P&N	-0.0034	0.0015	0.1035
	KR&N	KR	0.0123	0.0015	**0.0000**
		P	0.0118	0.0015	**0.0000**
		P&N	0.0084	0.0015	**0.0000**
	P&N	KR	0.0039	0.0015	0.0433
		P	0.0034	0.0015	0.1035
		KR&N	-0.0084	0.0015	**0.0000**

ตารางที่ 4.3 (ต่อ) ผลทดสอบความแตกต่างระหว่างค่าเฉลี่ยแต่ละคู่ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ I เมื่อ K = 3, ρ = 0.2 และ n = 20 ด้วยวิธี Tukey

ค่าที่วัด	ระยะห่าง		ผลต่างค่าเฉลี่ย	Std. Error	Sig.
ค่า Jaccard coefficient	KR	P	-0.0015	0.0038	0.9810
		KR&N	-0.0321	0.0038	**0.0000**
		P&N	-0.0102	0.0038	0.0396
	P	KR	0.0015	0.0038	0.9810
		KR&N	-0.0307	0.0038	**0.0000**
		P&N	-0.0087	0.0038	0.1043
	KR&N	KR	0.0321	0.0038	**0.0000**
		P	0.0307	0.0038	**0.0000**
		P&N	0.0219	0.0038	**0.0000**
	P&N	KR	0.0102	0.0038	0.0396
		P	0.0087	0.0038	0.1043
		KR&N	-0.0219	0.0038	**0.0000**
ค่า Average silhouette width	KR	P	-0.0063	0.0017	**0.0015**
		KR&N	0.0579	0.0017	**0.0000**
		P&N	0.0463	0.0017	**0.0000**
	P	KR	0.0063	0.0017	**0.0015**
		KR&N	0.0642	0.0017	**0.0000**
		P&N	0.0526	0.0017	**0.0000**
	KR&N	KR	-0.0579	0.0017	**0.0000**
		P	-0.0642	0.0017	**0.0000**
		P&N	-0.0116	0.0017	**0.0000**
	P&N	KR	-0.0463	0.0017	**0.0000**
		P	-0.0526	0.0017	**0.0000**
		KR&N	0.0116	0.0017	**0.0000**

ข้อมูลรูปแบบที่ I เมื่อ $K = 3,\ \rho = 0.2$ *และ* $n = 100$

ผลวิเคราะห์ความแปรปรวน พบว่าค่า Purity ค่า Rand statistic ค่า Jaccard coefficient และค่า Average silhouette width ต่างมี Sig. = 0.0000 ดังตารางที่ 4.4 ซึ่งน้อยกว่า 0.05 จึงปฏิเสธสมมติฐานหลัก นั่นคือประสิทธิภาพการวิเคราะห์กลุ่มขึ้นกับระยะห่างที่ใช้ในการวิเคราะห์กลุ่มที่ระดับนัยสำคัญ 0.05 และค่า Average silhouette width มีค่าขึ้นกับระยะห่างที่ใช้ในการวิเคราะห์กลุ่มที่ระดับนัยสำคัญ 0.05 เช่นเดียวกัน

ตารางที่ 4.4 ผลวิเคราะห์ความแปรปรวนของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ I เมื่อ $K = 3,\ \rho = 0.2$ และ $n = 100$

แหล่งความแปรปรวน		SS	df	MS	F	Sig.
ค่า Purity	ระหว่างกลุ่ม	0.1149	3	0.0383	264.6944	**0.0000**
	ภายในกลุ่ม	0.5782	3996	0.0001		
	ผลรวม	0.6932	3999			
ค่า Rand statistic	ระหว่างกลุ่ม	0.1828	3	0.0609	275.6753	**0.0000**
	ภายในกลุ่ม	0.8831	3996	0.0002		
	ผลรวม	1.0659	3999			
ค่า Jaccard coefficient	ระหว่างกลุ่ม	1.2244	3	0.4081	281.6805	**0.0000**
	ภายในกลุ่ม	5.7898	3996	0.0014		
	ผลรวม	7.0142	3999			
ค่า Average silhouette width	ระหว่างกลุ่ม	4.6870	3	1.5623	5563.3321	**0.0000**
	ภายในกลุ่ม	1.1222	3996	0.0003		
	ผลรวม	5.8092	3999			

ผู้วิจัยจึงทดสอบความแตกต่างของค่าเฉลี่ยของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแต่ละคู่ ด้วยวิธี Tukey พบว่าค่าเฉลี่ยของค่า Purity ค่า Rand statistic และค่า Jaccard coefficient ที่ได้จากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR และระยะห่างของ P ไม่แตกต่างกันเพียงคู่เดียว แสดงว่าการวิเคราะห์กลุ่มด้วยระยะห่างของ KR และระยะห่างของ P มีประสิทธิภาพไม่แตกต่างกัน ขณะที่การวิเคราะห์กลุ่มด้วยระยะห่างวิธีอื่น ๆ มีประสิทธิภาพโดยเฉลี่ยแตกต่างกันที่ระดับนัยสำคัญ 0.05 นอกจากนี้ค่า Average silhouette width จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ ทั้ง 4 วิธีนี้ แตกต่างกันที่ระดับนัยสำคัญ 0.05 ดังตารางที่ 4.5

ตารางที่ 4.5 ผลทดสอบความแตกต่างระหว่างค่าเฉลี่ยแต่ละคู่ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ I เมื่อ K = 3, ρ = 0.2 และ n = 100 ด้วยวิธี Tukey

ค่าที่วัด	ระยะห่าง		ผลต่างค่าเฉลี่ย	Std. Error	Sig.
ค่า Purity	KR	P	0.0000	0.0005	0.9997
		KR&N	-0.0131	0.0005	**0.0000**
		P&N	-0.0039	0.0005	**0.0000**
	P	KR	0.0000	0.0005	0.9997
		KR&N	-0.0131	0.0005	**0.0000**
		P&N	-0.0039	0.0005	**0.0000**
	KR&N	KR	0.0131	0.0005	**0.0000**
		P	0.0131	0.0005	**0.0000**
		P&N	0.0092	0.0005	**0.0000**
	P&N	KR	0.0039	0.0005	**0.0000**
		P	0.0039	0.0005	**0.0000**
		KR&N	-0.0092	0.0005	**0.0000**
ค่า Rand statistic	KR	P	0.0000	0.0007	1.0000
		KR&N	-0.0165	0.0007	**0.0000**
		P&N	-0.0049	0.0007	**0.0000**
	P	KR	0.0000	0.0007	1.0000
		KR&N	-0.0165	0.0007	**0.0000**
		P&N	-0.0049	0.0007	**0.0000**
	KR&N	KR	0.0165	0.0007	**0.0000**
		P	0.0165	0.0007	**0.0000**
		P&N	0.0116	0.0007	**0.0000**
	P&N	KR	0.0049	0.0007	**0.0000**
		P	0.0049	0.0007	**0.0000**
		KR&N	-0.0116	0.0007	**0.0000**

ตารางที่ 4.5 (ต่อ) ผลทดสอบความแตกต่างระหว่างค่าเฉลี่ยแต่ละคู่ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ I เมื่อ $K = 3$, $\rho = 0.2$ และ $n = 100$ ด้วยวิธี Tukey

ค่าที่วัด	ระยะห่าง		ผลต่างค่าเฉลี่ย	Std. Error	Sig.
ค่า Jaccard coefficient	KR	P	-0.0001	0.0017	1.0000
		KR&N	-0.0429	0.0017	**0.0000**
		P&N	-0.0126	0.0017	**0.0000**
	P	KR	0.0001	0.0017	1.0000
		KR&N	-0.0428	0.0017	**0.0000**
		P&N	-0.0125	0.0017	**0.0000**
	KR&N	KR	0.0429	0.0017	**0.0000**
		P	0.0428	0.0017	**0.0000**
		P&N	0.0303	0.0017	**0.0000**
	P&N	KR	0.0126	0.0017	**0.0000**
		P	0.0125	0.0017	**0.0000**
		KR&N	-0.0303	0.0017	**0.0000**
ค่า Average silhouette width	KR	P	-0.0060	0.0007	**0.0000**
		KR&N	0.0708	0.0007	**0.0000**
		P&N	0.0588	0.0007	**0.0000**
	P	KR	0.0060	0.0007	**0.0000**
		KR&N	0.0768	0.0007	**0.0000**
		P&N	0.0648	0.0007	**0.0000**
	KR&N	KR	-0.0708	0.0007	**0.0000**
		P	-0.0768	0.0007	**0.0000**
		P&N	-0.0121	0.0007	**0.0000**
	P&N	KR	-0.0588	0.0007	**0.0000**
		P	-0.0648	0.0007	**0.0000**
		KR&N	0.0121	0.0007	**0.0000**

ข้อมูลรูปแบบที่ I เมื่อ $K = 3,\ \rho = 0.8$ *และ* $n = 20$

ผลวิเคราะห์ความแปรปรวน พบว่าค่า Purity ค่า Rand statistic ค่า Jaccard coefficient และค่า Average silhouette width ต่างมี Sig. ดังตารางที่ 4.6 ซึ่งน้อยกว่า 0.05 จึงปฏิเสธสมมติฐานหลัก นั่นคือประสิทธิภาพการวิเคราะห์กลุ่มขึ้นกับระยะห่างที่ใช้ในการวิเคราะห์กลุ่มที่ระดับนัยสำคัญ 0.05 และค่า Average silhouette width มีค่าขึ้นกับระยะห่างที่ระดับนัยสำคัญ 0.05 เช่นเดียวกัน

ตารางที่ 4.6 ผลวิเคราะห์ความแปรปรวนของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ I เมื่อ $K = 3,\ \rho = 0.8$ และ $n = 20$

แหล่งความแปรปรวน		SS	df	MS	F	Sig.
ค่า Purity	ระหว่างกลุ่ม	0.0335	3	0.0112	5.0810	**0.0016**
	ภายในกลุ่ม	8.7802	3996	0.0022		
	ผลรวม	8.8136	3999			
ค่า Rand statistic	ระหว่างกลุ่ม	0.0837	3	0.0279	12.6332	**0.0000**
	ภายในกลุ่ม	8.8298	3996	0.0022		
	ผลรวม	8.9136	3999			
ค่า Jaccard coefficient	ระหว่างกลุ่ม	0.2547	3	0.0849	12.1355	**0.0000**
	ภายในกลุ่ม	27.9521	3996	0.0070		
	ผลรวม	28.2068	3999			
ค่า Average silhouette width	ระหว่างกลุ่ม	4.2958	3	1.4319	485.4619	**0.0000**
	ภายในกลุ่ม	11.7867	3996	0.0029		
	ผลรวม	16.0825	3999			

ผู้วิจัยทดสอบความแตกต่างของค่าเฉลี่ยของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแต่ละคู่ ด้วยวิธี Tukey พบว่าค่าเฉลี่ยของค่า Purity ที่ได้จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบ KR&N ต่างก็แตกต่างจากระยะห่างของ P และระยะห่างแบบ P&N ที่ระดับนัยสำคัญ 0.05 แต่ไม่แตกต่างจากระยะห่างของ KR ขณะที่ค่าเฉลี่ยของค่า Rand statistic และค่า Jaccard coefficient ที่ได้จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบ KR&N ต่างก็แตกต่างจากระยะห่างทั้ง 3 วิธีที่ระดับนัยสำคัญ 0.05 จึงกล่าวได้ว่าในภาพรวมการวิเคราะห์กลุ่มด้วยระยะห่างแบบ KR&N มีประสิทธิภาพแตกต่างจากการวิเคราะห์กลุ่มด้วยระยะห่างอีก 3 วิธี และการวิเคราะห์กลุ่มด้วยระยะห่างทั้ง 3 วิธีนี้มีประสิทธิภาพ

ไม่แตกต่างกัน นอกจากนี้ค่าเฉลี่ยของค่า Average silhouette width จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ ทั้ง 4 วิธีนี้ แตกต่างกันที่ระดับนัยสำคัญ 0.05 ดังตารางที่ 4.7

ตารางที่ 4.7 ผลทดสอบความแตกต่างระหว่างค่าเฉลี่ยแต่ละคู่ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ I เมื่อ K = 3, ρ = 0.8 และ n = 20 ด้วยวิธี Tukey

ค่าที่วัด	ระยะห่าง		ผลต่างค่าเฉลี่ย	Std. Error	Sig.
ค่า Purity	KR	P	0.0025	0.0021	0.6215
		KR&N	-0.0052	0.0021	0.0657
		P&N	0.0010	0.0021	0.9625
	P	KR	-0.0025	0.0021	0.6215
		KR&N	-0.0077	0.0021	**0.0014**
		P&N	-0.0015	0.0021	0.8877
	KR&N	KR	0.0052	0.0021	0.0657
		P	0.0077	0.0021	**0.0014**
		P&N	0.0062	0.0021	**0.0169**
	P&N	KR	-0.0010	0.0021	0.9625
		P	0.0015	0.0021	0.8877
		KR&N	-0.0062	0.0021	**0.0169**
ค่า Rand statistic	KR	P	0.0047	0.0021	0.1148
		KR&N	-0.0078	0.0021	**0.0012**
		P&N	0.0013	0.0021	0.9247
	P	KR	-0.0047	0.0021	0.1148
		KR&N	-0.0125	0.0021	**0.0000**
		P&N	-0.0034	0.0021	0.3736
	KR&N	KR	0.0078	0.0021	**0.0012**
		P	0.0125	0.0021	**0.0000**
		P&N	0.0091	0.0021	**0.0001**
	P&N	KR	-0.0013	0.0021	0.9247
		P	0.0034	0.0021	0.3736
		KR&N	-0.0091	0.0021	**0.0001**

ตารางที่ 4.7 (ต่อ) ผลทดสอบความแตกต่างระหว่างค่าเฉลี่ยแต่ละคู่ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ I เมื่อ K = 3, ρ = 0.8 และ n = 20 ด้วยวิธี Tukey

ค่าที่วัด	ระยะห่าง		ผลต่างค่าเฉลี่ย	Std. Error	Sig.
ค่า Jaccard coefficient	KR	P	0.0070	0.0037	0.2449
		KR&N	-0.0147	0.0037	**0.0005**
		P&N	0.0010	0.0037	0.9923
	P	KR	-0.0070	0.0037	0.2449
		KR&N	-0.0217	0.0037	**0.0000**
		P&N	-0.0059	0.0037	0.3895
	KR&N	KR	0.0147	0.0037	**0.0005**
		P	0.0217	0.0037	**0.0000**
		P&N	0.0158	0.0037	**0.0002**
	P&N	KR	-0.0010	0.0037	0.9923
		P	0.0059	0.0037	0.3895
		KR&N	-0.0158	0.0037	**0.0002**
ค่า Average silhouette width	KR	P	-0.0144	0.0024	**0.0000**
		KR&N	0.0670	0.0024	**0.0000**
		P&N	0.0439	0.0024	**0.0000**
	P	KR	0.0144	0.0024	**0.0000**
		KR&N	0.0814	0.0024	**0.0000**
		P&N	0.0583	0.0024	**0.0000**
	KR&N	KR	-0.0670	0.0024	**0.0000**
		P	-0.0814	0.0024	**0.0000**
		P&N	-0.0231	0.0024	**0.0000**
	P&N	KR	-0.0439	0.0024	**0.0000**
		P	-0.0583	0.0024	**0.0000**
		KR&N	0.0231	0.0024	**0.0000**

ข้อมูลรูปแบบที่ I เมื่อ $K = 3,\ \rho = 0.8$ *และ* $n = 100$

ผลวิเคราะห์ความแปรปรวน พบว่าค่า Purity ค่า Rand statistic ค่า Jaccard coefficient และค่า Average silhouette width ต่างมี Sig. = 0.0000 ดังตารางที่ 4.8 ซึ่งน้อยกว่า 0.05 จึงปฏิเสธสมมติฐานหลัก นั่นคือประสิทธิภาพการวิเคราะห์กลุ่มขึ้นกับระยะห่างที่ใช้ในการวิเคราะห์กลุ่มที่ระดับนัยสำคัญ 0.05 และค่า Average silhouette width ก็มีค่าขึ้นกับระยะห่างที่ระดับนัยสำคัญ 0.05 เช่นเดียวกัน

ตารางที่ 4.8 ผลวิเคราะห์ความแปรปรวนของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ I เมื่อ $K = 3,\ \rho = 0.8$ และ $n = 100$

แหล่งความแปรปรวน		SS	df	MS	F	Sig.
ค่า Purity	ระหว่างกลุ่ม	0.0504	3	0.0168	31.9618	**0.0000**
	ภายในกลุ่ม	2.0989	3996	0.0005		
	ผลรวม	2.1493	3999			
ค่า Rand statistic	ระหว่างกลุ่ม	0.1154	3	0.0385	72.2261	**0.0000**
	ภายในกลุ่ม	2.1288	3996	0.0005		
	ผลรวม	2.2442	3999			
ค่า Jaccard coefficient	ระหว่างกลุ่ม	0.3200	3	0.1067	66.3262	**0.0000**
	ภายในกลุ่ม	6.4268	3996	0.0016		
	ผลรวม	6.7468	3999			
ค่า Average silhouette width	ระหว่างกลุ่ม	6.9222	3	2.3074	3948.1313	**0.0000**
	ภายในกลุ่ม	2.3354	3996	0.0006		
	ผลรวม	9.2575	3999			

ผู้วิจัยทดสอบความแตกต่างของค่าเฉลี่ยของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแต่ละคู่ ด้วยวิธี Tukey พบว่าค่าเฉลี่ยของค่า Purity ค่า Rand statistic และค่า Jaccard coefficient ที่ได้จากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR และระยะห่างแบบ P&N ไม่แตกต่างกันเพียงคู่เดียว แสดงว่าโดยเฉลี่ยการวิเคราะห์กลุ่มด้วยระยะห่างของ KR และระยะห่างแบบ P&N มีประสิทธิภาพไม่แตกต่างกัน ขณะที่การวิเคราะห์กลุ่มด้วยระยะห่างคู่อื่น ๆ มีประสิทธิภาพแตกต่างกันที่ระดับนัยสำคัญ 0.05 นอกจากนี้ค่า Average silhouette width จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ ทั้ง 4 วิธีนี้ แตกต่างกันที่ระดับนัยสำคัญ 0.05 ดังตารางที่ 4.9

ตารางที่ 4.9 ผลทดสอบความแตกต่างระหว่างค่าเฉลี่ยแต่ละคู่ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ I เมื่อ K = 3, ρ = 0.8 และ n = 100 ด้วยวิธี Tukey

ค่าที่วัด	ระยะห่าง		ผลต่างค่าเฉลี่ย	Std. Error	Sig.
ค่า Purity	KR	P	0.0040	0.0010	**0.0006**
		KR&N	-0.0057	0.0010	**0.0000**
		P&N	0.0013	0.0010	0.5749
	P	KR	-0.0040	0.0010	**0.0006**
		KR&N	-0.0097	0.0010	**0.0000**
		P&N	-0.0027	0.0010	**0.0436**
	KR&N	KR	0.0057	0.0010	**0.0000**
		P	0.0097	0.0010	**0.0000**
		P&N	0.0070	0.0010	**0.0000**
	P&N	KR	-0.0013	0.0010	0.5749
		P	0.0027	0.0010	**0.0436**
		KR&N	-0.0070	0.0010	**0.0000**
ค่า Rand statistic	KR	P	0.0066	0.0010	**0.0000**
		KR&N	-0.0081	0.0010	**0.0000**
		P&N	0.0026	0.0010	0.0622
	P	KR	-0.0066	0.0010	**0.0000**
		KR&N	-0.0147	0.0010	**0.0000**
		P&N	-0.0041	0.0010	**0.0005**
	KR&N	KR	0.0081	0.0010	**0.0000**
		P	0.0147	0.0010	**0.0000**
		P&N	0.0106	0.0010	**0.0000**
	P&N	KR	-0.0026	0.0010	0.0622
		P	0.0041	0.0010	**0.0005**
		KR&N	-0.0106	0.0010	**0.0000**

ตารางที่ 4.9 (ต่อ) ผลทดสอบความแตกต่างระหว่างค่าเฉลี่ยแต่ละคู่ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ I เมื่อ $K = 3$, $\rho = 0.8$ และ $n = 100$ ด้วยวิธี Tukey

ค่าที่วัด	ระยะห่าง		ผลต่างค่าเฉลี่ย	Std. Error	Sig.
ค่า Jaccard coefficient	KR	P	0.0103	0.0018	**0.0000**
		KR&N	-0.0142	0.0018	**0.0000**
		P&N	0.0035	0.0018	0.2122
	P	KR	-0.0103	0.0018	**0.0000**
		KR&N	-0.0245	0.0018	**0.0000**
		P&N	-0.0068	0.0018	**0.0009**
	KR&N	KR	0.0142	0.0018	**0.0000**
		P	0.0245	0.0018	**0.0000**
		P&N	0.0177	0.0018	**0.0000**
	P&N	KR	-0.0035	0.0018	0.2122
		P	0.0068	0.0018	**0.0009**
		KR&N	-0.0177	0.0018	**0.0000**
ค่า Average silhouette width	KR	P	-0.0167	0.0011	**0.0000**
		KR&N	0.0856	0.0011	**0.0000**
		P&N	0.0575	0.0011	**0.0000**
	P	KR	0.0167	0.0011	**0.0000**
		KR&N	0.1023	0.0011	**0.0000**
		P&N	0.0742	0.0011	**0.0000**
	KR&N	KR	-0.0856	0.0011	**0.0000**
		P	-0.1023	0.0011	**0.0000**
		P&N	-0.0281	0.0011	**0.0000**
	P&N	KR	-0.0575	0.0011	**0.0000**
		P	-0.0742	0.0011	**0.0000**
		KR&N	0.0281	0.0011	**0.0000**

4.1.1.2 ผลทดสอบความแตกต่างการวิเคราะห์กลุ่มข้อมูลรูปแบบที่ I เมื่อ $K = 5$

เนื่องจากเมื่อ $K = 5$ ยังสามารถแบ่งข้อมูลได้อีกตามความแตกต่างของค่าสัมประสิทธิ์สหสัมพันธ์ และขนาดข้อมูลต่อกลุ่ม จึงทำการศึกษาประสิทธิภาพการวิเคราะห์กลุ่มข้อมูลต่าง ๆ ได้ดังต่อไปนี้

ข้อมูลรูปแบบที่ I เมื่อ $K = 5$, $\rho = 0.2$ และ $n = 20$

ผลวิเคราะห์ความแปรปรวน ดังตารางที่ 4.10 พบว่าค่า Purity ค่า Rand statistic และค่า Jaccard coefficient มี Sig. มากกว่า 0.05 จึงไม่สามารถปฏิเสธสมมติฐานหลัก ดังนั้นประสิทธิภาพการวิเคราะห์กลุ่มไม่ขึ้นกับระยะห่างที่ใช้ในการวิเคราะห์กลุ่ม อย่างไรก็ตามค่า Average silhouette width มี Sig. = 0.0105 ซึ่งน้อยกว่า 0.05 จึงปฏิเสธสมมติฐานหลัก แสดงว่าค่า Average silhouette width มีค่าขึ้นกับระยะห่างที่ใช้ในการวิเคราะห์กลุ่มที่ระดับนัยสำคัญ 0.05

ตารางที่ 4.10 ผลวิเคราะห์ความแปรปรวนของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ I เมื่อ $K = 5$, $\rho = 0.2$ และ $n = 20$

แหล่งความแปรปรวน		SS	df	MS	F	Sig.
ค่า Purity	ระหว่างกลุ่ม	0.0001	3	0.0000	0.0560	0.9826
	ภายในกลุ่ม	2.9106	3996	0.0007		
	ผลรวม	2.9107	3999			
ค่า Rand statistic	ระหว่างกลุ่ม	0.0001	3	0.0000	0.0490	0.9857
	ภายในกลุ่ม	1.4528	3996	0.0004		
	ผลรวม	1.4529	3999			
ค่า Jaccard coefficient	ระหว่างกลุ่ม	0.0008	3	0.0003	0.0465	0.9867
	ภายในกลุ่ม	22.1389	3996	0.0055		
	ผลรวม	22.1396	3999			
ค่า Average silhouette width	ระหว่างกลุ่ม	0.0133	3	0.0044	3.7508	**0.0105**
	ภายในกลุ่ม	4.7118	3996	0.0012		
	ผลรวม	4.7251	3999			

จากผลวิเคราะห์ความแปรปรวน ผู้วิจัยจึงทดสอบความแตกต่างระหว่างค่าเฉลี่ยแต่ละคู่ ของค่า Average silhouette width จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ ด้วยวิธี Tukey พบว่า ความแตกต่างของค่า Average silhouette width ที่ได้จากการวิเคราะห์กลุ่มระหว่างระยะห่างแต่ละคู่จากทั้ง 4 วิธีนี้ มี .Sig มากกว่าหรือเท่ากับ 0.05 ดังตารางที่ 4.11 จึงไม่สามารถปฏิเสธสมมติฐานหลักได้ แสดงว่าค่า Average silhouette width ที่ได้จากการวิเคราะห์กลุ่มระหว่างระยะห่างแต่ละคู่จากทั้ง 4 วิธีนี้ ไม่แตกต่างกัน ซึ่งขัดแย้งกับผลวิเคราะห์ความแปรปรวนก่อนหน้านี้

ตารางที่ 4.11 ผลทดสอบความแตกต่างระหว่างค่าเฉลี่ยแต่ละคู่ของค่า Average silhouette width จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ I เมื่อ K = 5, ρ = 0.2 และ n = 20 ด้วยวิธี Tukey

ระยะห่าง		ผลต่างค่าเฉลี่ย	Std. Error	Sig.
KR	P	-0.0003	0.0015	0.9978
	KR&N	0.0036	0.0015	0.0834
	P&N	0.0033	0.0015	0.1294
P	KR	0.0003	0.0015	0.9978
	KR&N	0.0039	0.0015	0.0526
	P&N	0.0036	0.0015	0.0849
KR&N	KR	-0.0036	0.0015	0.0834
	P	-0.0039	0.0015	0.0526
	P&N	-0.0003	0.0015	0.9976
P&N	KR	-0.0033	0.0015	0.1294
	P	-0.0036	0.0015	0.0849
	KR&N	0.0003	0.0015	0.9976

ข้อมูลรูปแบบที่ I เมื่อ $K = 5$, $\rho = 0.2$ และ $n = 100$

จากตารางที่ 4.12 และ 4.13 พบว่าสามารถสรุปผลทดสอบความแตกต่างประสิทธิภาพการวิเคราะห์กลุ่มข้อมูลได้เช่นเดียวกับกรณีข้อมูลรูปแบบที่ I เมื่อ $K = 5$, $\rho = 0.2$ และ $n = 20$

ตารางที่ 4.12 ผลวิเคราะห์ความแปรปรวนของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ I เมื่อ $K = 5$, $\rho = 0.2$ และ $n = 100$

แหล่งความแปรปรวน		SS	df	MS	F	Sig.
ค่า Purity	ระหว่างกลุ่ม	0.0000	3	0.0000	0.3210	0.8100
	ภายในกลุ่ม	0.5420	3996	0.0000		
	ผลรวม	0.5420	3999			
ค่า Rand statistic	ระหว่างกลุ่ม	0.0000	3	0.0000	0.3180	0.8130
	ภายในกลุ่ม	0.2770	3996	0.0000		
	ผลรวม	0.2770	3999			
ค่า Jaccard coefficient	ระหว่างกลุ่ม	0.0010	3	0.0000	0.3410	0.7960
	ภายในกลุ่ม	4.0700	3996	0.0010		
	ผลรวม	4.0710	3999			
ค่า Average silhouette width	ระหว่างกลุ่ม	0.0020	3	0.0010	3.2980	**0.0200**
	ภายในกลุ่ม	0.8680	3996	0.0000		
	ผลรวม	0.8700	3999			

ตารางที่ 4.13 ผลทดสอบความแตกต่างระหว่างค่าเฉลี่ยแต่ละคู่ของค่า Average silhouette width จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ I เมื่อ K = 5, ρ = 0.2 และ n = 100 ด้วยวิธี Tukey

ระยะห่าง		ผลต่างค่าเฉลี่ย	Std. Error	Sig.
KR	P	-0.0001	0.0007	0.9996
	KR&N	0.0015	0.0007	0.1178
	P&N	0.0014	0.0007	0.1462
P	KR	0.0001	0.0007	0.9996
	KR&N	0.0015	0.0007	0.0933
	P&N	0.0015	0.0007	0.1172
KR&N	KR	-0.0015	0.0007	0.1178
	P	-0.0015	0.0007	0.0933
	P&N	-0.0001	0.0007	0.9997
P&N	KR	-0.0014	0.0007	0.1462
	P	-0.0015	0.0007	0.1172
	KR&N	0.0001	0.0007	0.9997

ข้อมูลรูปแบบที่ I เมื่อ K = 5, ρ = 0.8 และ n = 20

ผลวิเคราะห์ความแปรปรวน ดังตารางที่ 4.14 พบว่าค่า Purity ค่า Rand statistic และค่า Jaccard coefficient มี Sig. มากกว่า 0.05 จึงไม่สามารถปฏิเสธสมมติฐานหลัก ดังนั้น ประสิทธิภาพการวิเคราะห์กลุ่มไม่ขึ้นกับระยะห่างที่ใช้ในการวิเคราะห์กลุ่ม อย่างไรก็ตามค่า Average silhouette width มี Sig. = 0.0000 ซึ่งน้อยกว่า 0.05 จึงปฏิเสธสมมติฐานหลัก แสดงว่าค่า Average silhouette width มีค่าขึ้นกับระยะห่างที่ใช้ในการวิเคราะห์กลุ่มที่ระดับนัยสำคัญ 0.05

ตารางที่ 4.14 ผลวิเคราะห์ความแปรปรวนของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ I เมื่อ $K = 5$, $\rho = 0.8$ และ $n = 20$

แหล่งความแปรปรวน		SS	df	MS	F	Sig.
ค่า Purity	ระหว่างกลุ่ม	0.0003	3	0.0001	0.0474	0.9863
	ภายในกลุ่ม	7.2190	3996	0.0018		
	ผลรวม	7.2192	3999			
ค่า Rand statistic	ระหว่างกลุ่ม	0.0000	3	0.0000	0.0019	0.9999
	ภายในกลุ่ม	2.2479	3996	0.0006		
	ผลรวม	2.2479	3999			
ค่า Jaccard coefficient	ระหว่างกลุ่ม	0.0005	3	0.0002	0.0385	0.9899
	ภายในกลุ่ม	15.7311	3996	0.0039		
	ผลรวม	15.7316	3999			
ค่า Average silhouette width	ระหว่างกลุ่ม	0.0443	3	0.0148	7.5508	0.0000
	ภายในกลุ่ม	7.8066	3996	0.0020		
	ผลรวม	7.8508	3999			

ผู้วิจัยทดสอบความแตกต่างระหว่างค่าเฉลี่ยแต่ละคู่ ของค่า Average silhouette width จากการวิเคราะห์กลุ่มด้วยระยะห่างต่าง ๆ ด้วยวิธี Tukey ดังตารางที่ 4.15 พบว่า ค่า Average silhouette width ที่ได้จากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR แตกต่างจากระยะห่างแบบ KR&N และระยะห่างแบบ P&N ที่ระดับนัยสำคัญ 0.05 แต่ไม่แตกต่างจากระยะห่างของ P ส่วนค่า Average silhouette width ที่ได้จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบ KR&N แตกต่างจากระยะห่างของ KR และระยะห่างของ P ที่ระดับนัยสำคัญ 0.05 แต่ไม่แตกต่างจากระยะห่างแบบ P&N

ตารางที่ 4.15 ผลทดสอบความแตกต่างระหว่างค่าเฉลี่ยแต่ละคู่ของค่า Average silhouette width จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ I เมื่อ K = 5, ρ = 0.8 และ n = 20 ด้วยวิธี Tukey

ระยะห่าง		ผลต่างค่าเฉลี่ย	Std. Error	Sig.
KR	P	0.0000	0.0020	1.0000
	KR&N	0.0068	0.0020	**0.0031**
	P&N	0.0064	0.0020	**0.0064**
P	KR	0.0000	0.0020	1.0000
	KR&N	0.0069	0.0020	**0.0029**
	P&N	0.0065	0.0020	**0.0059**
KR&N	KR	-0.0068	0.0020	**0.0031**
	P	-0.0069	0.0020	**0.0029**
	P&N	-0.0004	0.0020	0.9971
P&N	KR	-0.0064	0.0020	**0.0064**
	P	-0.0065	0.0020	**0.0059**
	KR&N	0.0004	0.0020	0.9971

ข้อมูลรูปแบบที่ I เมื่อ $K = 5,\ \rho = 0.8$ *และ* $n = 100$

จากตารางที่ 4.16 และ 4.17 พบว่าสามารถสรุปผลการวิเคราะห์กลุ่มข้อมูลได้เช่นเดียวกับกรณีข้อมูลรูปแบบที่ I เมื่อ $K = 5,\ \rho = 0.8$ และ $n = 20$

ตารางที่ 4.16 ผลวิเคราะห์ความแปรปรวนของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ I เมื่อ $K = 5,\ \rho = 0.8$ และ $n = 100$

แหล่งความแปรปรวน		SS	df	MS	F	Sig.
ค่า Purity	ระหว่างกลุ่ม	0.0007	3	0.0002	0.7259	0.5364
	ภายในกลุ่ม	1.2948	3996	0.0003		
	ผลรวม	1.2955	3999			
ค่า Rand statistic	ระหว่างกลุ่ม	0.0002	3	0.0001	0.6580	0.5779
	ภายในกลุ่ม	0.4162	3996	0.0001		
	ผลรวม	0.4164	3999			
ค่า Jaccard coefficient	ระหว่างกลุ่ม	0.0016	3	0.0005	0.7389	0.5287
	ภายในกลุ่ม	2.8380	3996	0.0007		
	ผลรวม	2.8396	3999			
ค่า Average silhouette width	ระหว่างกลุ่ม	0.0073	3	0.0024	6.7386	0.0002
	ภายในกลุ่ม	1.4381	3996	0.0004		
	ผลรวม	1.4453	3999			

ตารางที่ 4.17 ผลทดสอบความแตกต่างระหว่างค่าเฉลี่ยแต่ละคู่ของค่า Average silhouette width จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ I เมื่อ K = 5, ρ = 0.8 และ n = 100 ด้วยวิธี Tukey

ระยะห่าง		ผลต่างค่าเฉลี่ย	Std. Error	Sig.
KR	P	0.0000	0.0008	1.0000
	KR&N	0.0027	0.0008	**0.0068**
	P&N	0.0027	0.0008	**0.0096**
P	KR	0.0000	0.0008	1.0000
	KR&N	0.0027	0.0008	**0.0069**
	P&N	0.0027	0.0008	**0.0097**
KR&N	KR	-0.0027	0.0008	**0.0068**
	P	-0.0027	0.0008	**0.0069**
	P&N	-0.0001	0.0008	0.9996
P&N	KR	-0.0027	0.0008	**0.0096**
	P	-0.0027	0.0008	**0.0097**
	KR&N	0.0001	0.0008	0.9996

4.1.2 ผลการวิเคราะห์กลุ่มข้อมูลรูปแบบที่ II

ผู้วิจัยทำการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ 4 วิธี ได้แก่ ระยะห่างของ KR ระยะห่างของ P ระยะห่างแบบ KR&N และระยะห่างแบบ P&N สำหรับข้อมูลรูปแบบที่ II ซึ่งเป็นข้อมูลที่ประกอบไปด้วยตัวแปรนามบัญญัติ และตัวแปรอันดับ อย่างละ 3 ตัวแปร และคำนวณค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของ ค่า Purity ค่า Rand statistic ค่า Jaccard coefficient และค่า Average silhouette width ได้ผลดังตารางที่ 4.18

ผลการเปรียบเทียบประสิทธิภาพในการวิเคราะห์กลุ่มด้วยมาตรวัดระยะห่างแบบต่าง ๆ โดยเฉลี่ย ดังตารางที่ 4.2 พบว่า

เมื่อ K = 3 และ ρ = 0.2 การวิเคราะห์กลุ่มด้วยระยะห่างของ P ให้ค่าเฉลี่ยของ ค่า Purity ค่า Rand statistic และค่า Jaccard coefficient สูงที่สุด ทุกกรณี แสดงว่าการวิเคราะห์กลุ่มด้วยระยะห่างของ P มีประสิทธิภาพดีที่สุดโดยเฉลี่ย

เมื่อ K = 3 และ ρ = 0.8 การวิเคราะห์กลุ่มด้วยระยะห่างของ KR ให้ค่าเฉลี่ยของ ค่า Purity ค่า Rand statistic และค่า Jaccard coefficient สูงที่สุด ทุกกรณี แสดงว่าการวิเคราะห์กลุ่มด้วยระยะห่างของ KR มีประสิทธิภาพดีที่สุดโดยเฉลี่ย

เมื่อ K = 5 และ ρ = 0.2 การวิเคราะห์กลุ่มด้วยระยะห่างของ KR ให้ค่าเฉลี่ยของ ค่า Purity ค่า Rand statistic และค่า Jaccard coefficient สูงที่สุด ทุกกรณี แสดงว่าการวิเคราะห์กลุ่มด้วยระยะห่างของ KR มีประสิทธิภาพดีที่สุดโดยเฉลี่ย

เมื่อ K = 5 ρ = 0.8 และ n = 20 การวิเคราะห์กลุ่มด้วยระยะห่างของ P ให้ค่าเฉลี่ยของค่า Purity ค่า Rand statistic และค่า Jaccard coefficient สูงที่สุด แสดงว่าการวิเคราะห์กลุ่มด้วยระยะห่างของ P มีประสิทธิภาพดีที่สุดโดยเฉลี่ย แต่เมื่อ n = 100 การวิเคราะห์กลุ่มด้วยระยะห่างของ P&N ให้ค่าเฉลี่ยของ ค่า Purity ค่า Rand statistic และค่า Jaccard coefficient สูงที่สุด แสดงว่าการวิเคราะห์กลุ่มด้วยระยะห่างของ P&N มีประสิทธิภาพดีที่สุดโดยเฉลี่ย

นอกจากนี้ เมื่อพิจารณาค่าเฉลี่ยของค่า Average silhouette width จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ ดังตารางที่ 4.18 พบว่า เมื่อ K = 3 ค่าเฉลี่ยของค่า Average silhouette width จากการวิเคราะห์กลุ่มด้วยระยะห่างของ P มีค่าสูงที่สุดในทุกกรณี ขณะที่เมื่อ K = 5 ค่าเฉลี่ยของค่า Average silhouette width จากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR และระยะห่างของ P มีค่าเท่ากันหรือใกล้เคียงกัน และสูงที่สุดในทุกกรณี

ตารางที่ 4.18 ค่าเฉลี่ย (ส่วนเบี่ยงเบนมาตรฐาน) จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ II

K	ρ	n	ค่าที่ได้	ระยะห่าง			
				KR	P	KR&N	P&N
3	0.2	20	ค่า Purity	0.9195 (0.0358)	**0.9233** (0.0356)	0.9085 (0.0385)	0.9036 (0.0386)
			ค่า Rand statistic	0.8997 (0.0422)	**0.9039** (0.0422)	0.8863 (0.0448)	0.8805 (0.0445)
			ค่า Jaccard coefficient	0.7379 (0.0963)	**0.7481** (0.0968)	0.7093 (0.0970)	0.6965 (0.0948)
			ค่า Average silhouette width	0.4158 (0.0473)	**0.4238** (0.0474)	0.3123 (0.0408)	0.3386 (0.0435)
		100	ค่า Purity	0.9178 (0.0151)	**0.9213** (0.0153)	0.9114 (0.0155)	0.9024 (0.0160)
			ค่า Rand statistic	0.8969 (0.0177)	**0.9010** (0.0181)	0.8889 (0.0181)	0.8780 (0.0186)
			ค่า Jaccard coefficient	0.7325 (0.0398)	**0.7420** (0.0408)	0.7155 (0.0392)	0.6925 (0.0391)
			ค่า Average silhouette width	0.4161 (0.0205)	**0.4238** (0.0204)	0.2993 (0.0169)	0.3269 (0.0182)
	0.8	20	ค่า Purity	**0.8107** (0.0507)	0.8094 (0.0495)	0.8008 (0.0507)	0.8003 (0.0486)
			ค่า Rand statistic	**0.7810** (0.0512)	0.7790 (0.0499)	0.7660 (0.0567)	0.7691 (0.0491)
			ค่า Jaccard coefficient	**0.5088** (0.0844)	0.5060 (0.0816)	0.4889 (0.0834)	0.4908 (0.0769)
			ค่า Average silhouette width	0.5346 (0.0570)	**0.5502** (0.0569)	0.4375 (0.0519)	0.4753 (0.0543)
		100	ค่า Purity	**0.8111** (0.0221)	0.8090 (0.0227)	0.8045 (0.0230)	0.8011 (0.0226)
			ค่า Rand statistic	**0.7802** (0.0220)	0.7775 (0.0226)	0.7725 (0.0234)	0.7692 (0.0225)
			ค่า Jaccard coefficient	**0.5106** (0.0353)	0.5072 (0.0359)	0.4998 (0.0359)	0.4947 (0.0348)
			ค่า Average silhouette width	0.5367 (0.0261)	**0.5539** (0.0256)	0.4126 (0.0231)	0.4590 (0.0245)

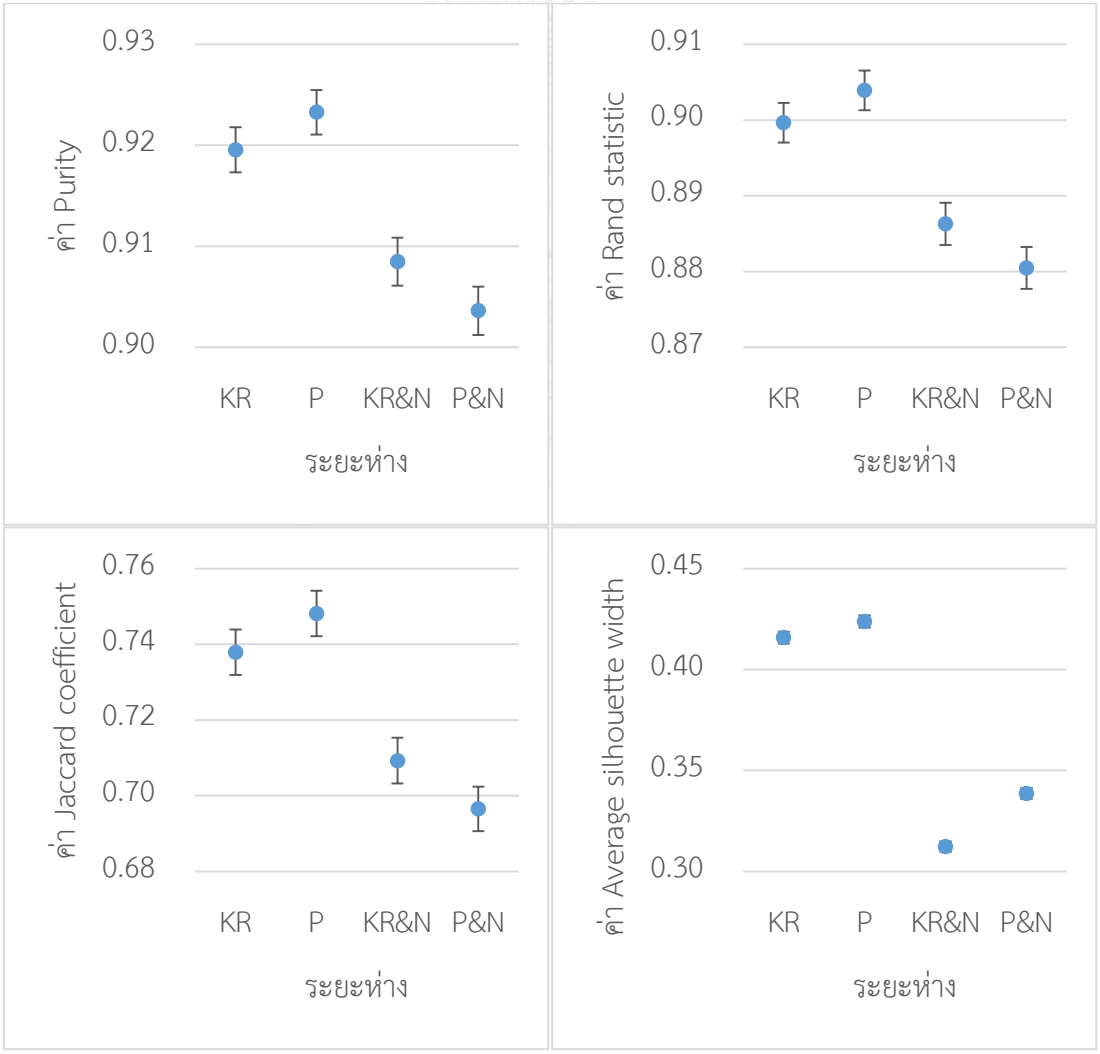
ตารางที่ 4.18 (ต่อ) ค่าเฉลี่ย (ส่วนเบี่ยงเบนมาตรฐาน) จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ II

K	ρ	n	ค่าที่ได้	ระยะห่าง			
				KR	P	KR&N	P&N
5	0.2	20	ค่า Purity	**0.8720** (0.0333)	0.8710 (0.0332)	0.8688 (0.0343)	0.8691 (0.0344)
			ค่า Rand statistic	**0.9076** (0.0218)	0.9069 (0.0218)	0.9053 (0.0226)	0.9055 (0.0226)
			ค่า Jaccard coefficient	**0.6174** (0.0737)	0.6152 (0.0731)	0.6105 (0.0750)	0.6111 (0.0752)
			ค่า Average silhouette width	**0.3897** (0.0372)	**0.3897** (0.0372)	0.3836 (0.0373)	0.3837 (0.0373)
		100	ค่า Purity	**0.8748** (0.0150)	0.8731 (0.0151)	0.8737 (0.0151)	0.8736 (0.0151)
			ค่า Rand statistic	**0.9083** (0.0100)	0.9072 (0.0100)	0.9076 (0.0100)	0.9075 (0.0101)
			ค่า Jaccard coefficient	**0.6257** (0.0331)	0.6221 (0.0330)	0.6233 (0.0332)	0.6231 (0.0332)
			ค่า Average silhouette width	**0.3931** (0.0160)	0.3930 (0.0161)	0.3912 (0.0161)	0.3913 (0.0161)
	0.8	20	ค่า Purity	0.7459 (0.0433)	**0.7461** (0.0433)	0.7455 (0.0433)	0.7459 (0.0432)
			ค่า Rand statistic	0.8326 (0.0231)	**0.8327** (0.0230)	0.8320 (0.0231)	0.8323 (0.0231)
			ค่า Jaccard coefficient	0.3999 (0.0578)	**0.4001** (0.0577)	0.3991 (0.0576)	0.3996 (0.0575)
			ค่า Average silhouette width	**0.5759** (0.0445)	0.5758 (0.0445)	0.5712 (0.0452)	0.5714 (0.0453)
		100	ค่า Purity	0.7461 (0.0194)	0.7467 (0.0194)	0.7467 (0.0193)	**0.7468** (0.0194)
			ค่า Rand statistic	0.8308 (0.0105)	**0.8311** (0.0105)	0.8310 (0.0104)	**0.8311** (0.0104)
			ค่า Jaccard coefficient	0.4035 (0.0259)	0.4043 (0.0259)	0.4042 (0.0258)	**0.4044** (0.0259)
			ค่า Average silhouette width	**0.5779** (0.0204)	**0.5779** (0.0205)	0.5760 (0.0205)	0.5761 (0.0205)

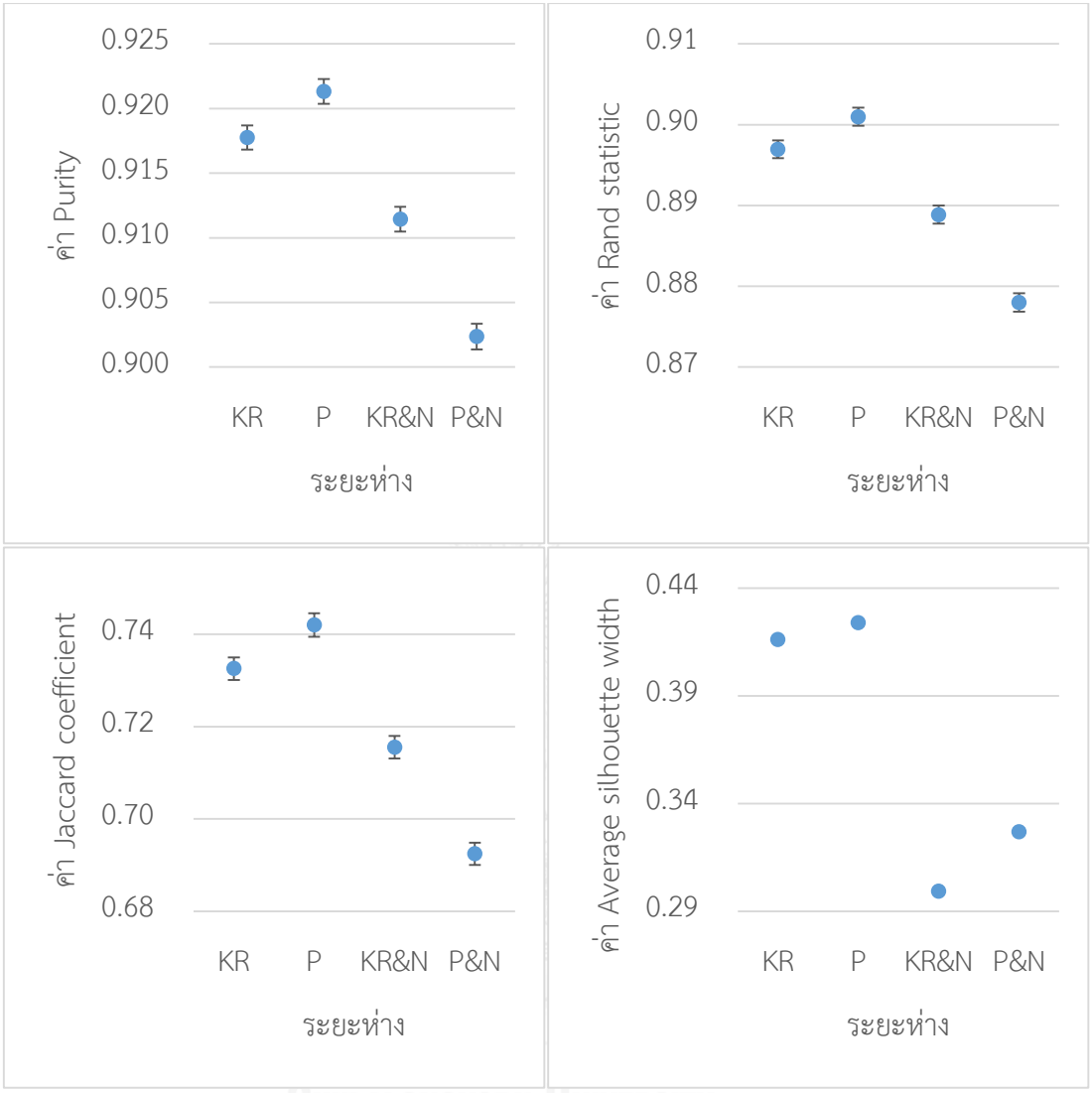
พิจารณากราฟช่วงความเชื่อมั่น 95% ของค่าต่าง ๆ พบว่า

เมื่อ K = 3 ค่า Average silhouette width จากการวิเคราะห์กลุ่มด้วยระยะห่างของ P ระยะห่างของ KR ระยะห่างแบบ P&N และระยะห่างแบบ KR&N มีค่าจากมากไปน้อยตามลำดับ และไม่มีส่วนของช่วงความเชื่อมั่น 95% ที่ทับซ้อนกันในทุกกรณี ดังรูปที่ 4.9 ถึง 4.12

เมื่อ K = 3 และ ρ = 0.2 ค่าเฉลี่ยของค่า Purity ค่า Rand statistic และค่า Jaccard coefficient จากการวิเคราะห์กลุ่มด้วยระยะห่างของ P มีค่ามากที่สุด ตามด้วยระยะห่างของ KR ระยะห่างแบบ KR&N และระยะห่างแบบ P&N โดยมีส่วนของช่วงความเชื่อมั่น 95% ทับซ้อนกันบางส่วน ดังรูปที่ 4.9 และ 4.10 แสดงว่าการวิเคราะห์กลุ่มด้วยระยะห่างของ P ระยะห่างของ KR ระยะห่างแบบ KR&N และระยะห่างแบบ P&N มีประสิทธิภาพดีที่สุดโดยเฉลี่ยตามลำดับ

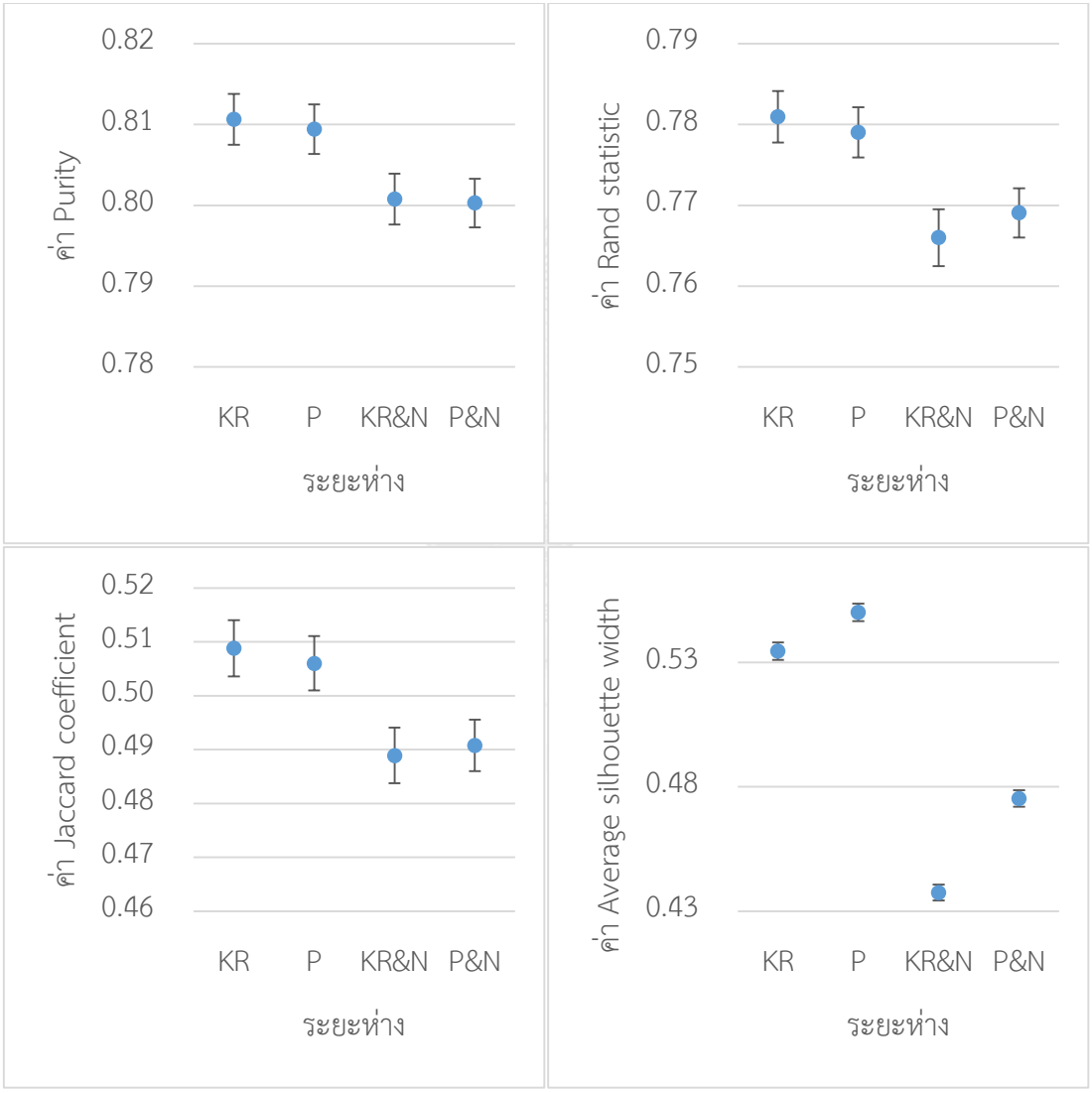


รูปที่ 4.9 กราฟช่วงความเชื่อมั่น 95% ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ II เมื่อ K = 3, ρ = 0.2 และ n = 20



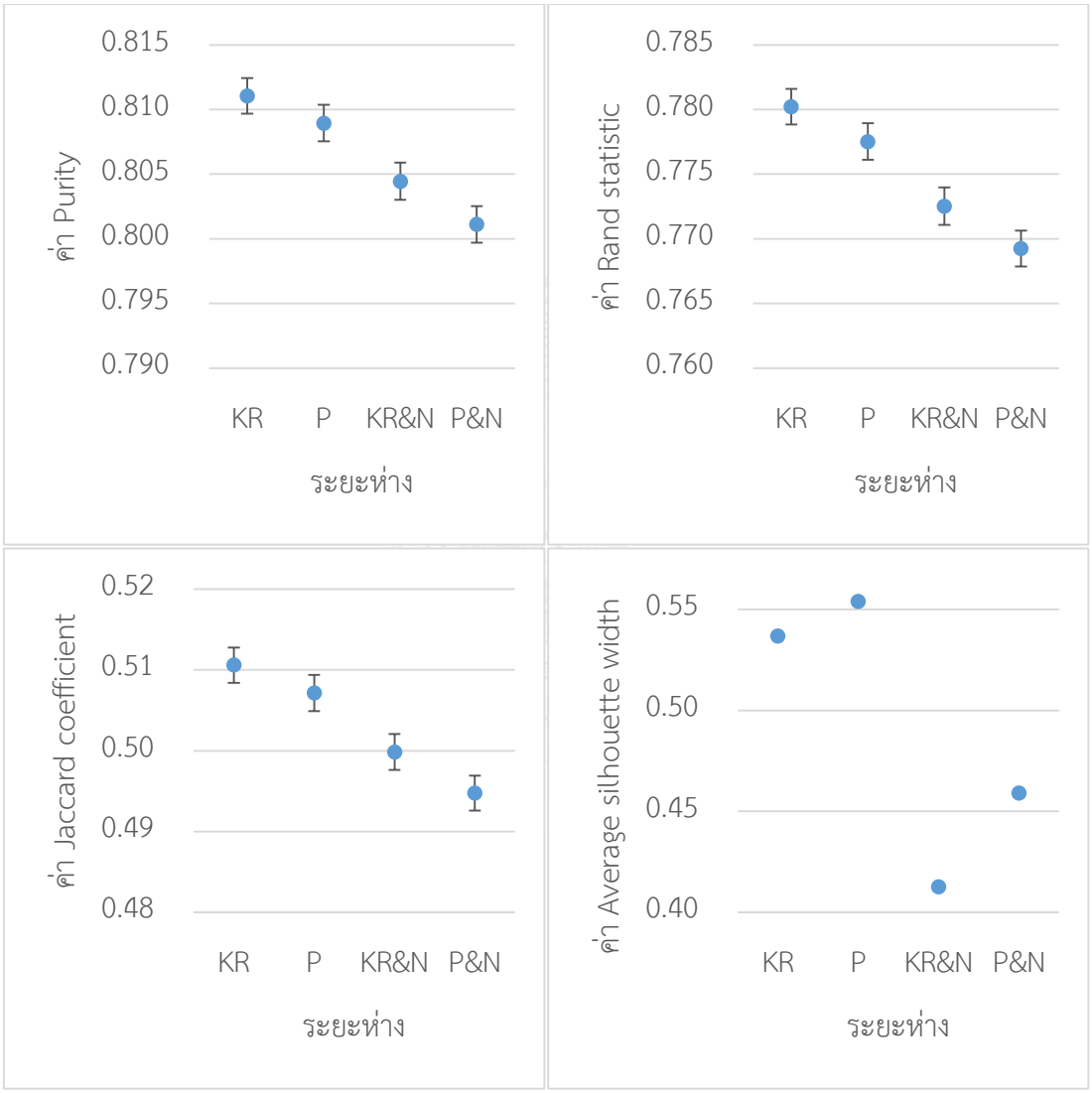
รูปที่ 4.10 กราฟช่วงความเชื่อมั่น 95% ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ II เมื่อ $K = 3$, $\rho = 0.2$ และ $n = 100$

เมื่อ $K = 3$ $\rho = 0.8$ และ $n = 20$ แม้ว่าการวิเคราะห์กลุ่มด้วยระยะห่างของ KR มีประสิทธิภาพดีที่สุดโดยเฉลี่ย แต่เมื่อพิจารณาค่าเฉลี่ยของค่า Purity ค่า Rand statistic และค่า Jaccard coefficient ดังรูปที่ 4.11 พบว่าการวิเคราะห์กลุ่มด้วยระยะห่างของ KR และระยะห่างของ P มีประสิทธิภาพโดยเฉลี่ยใกล้เคียงกัน และดีกว่าการวิเคราะห์กลุ่มด้วยระยะห่างแบบ KR&N และระยะห่างแบบ P&N ซึ่งมีค่าใกล้เคียงกันเช่นกัน



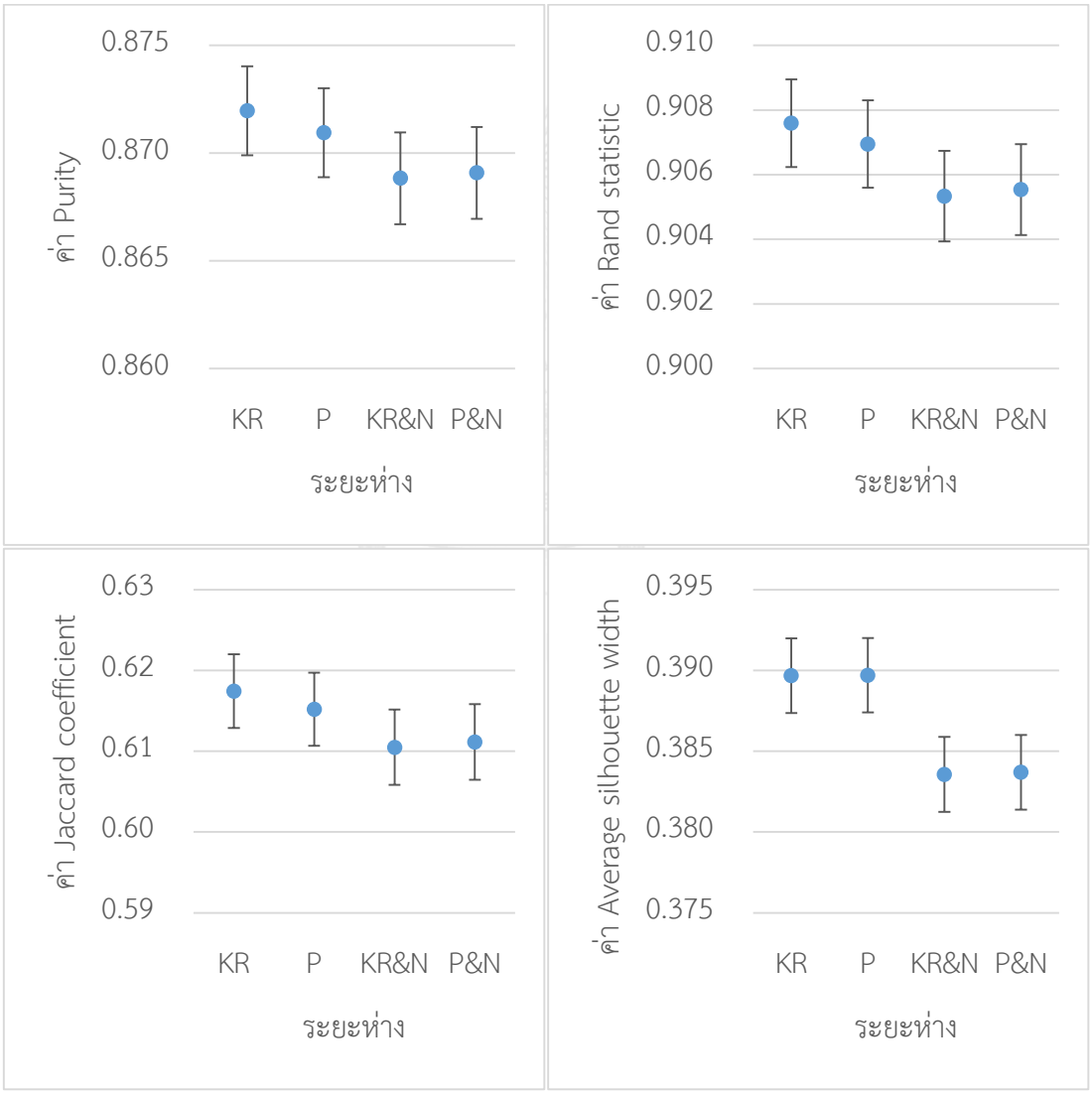
รูปที่ 4.11 กราฟช่วงความเชื่อมั่น 95% ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ II เมื่อ $K = 3$, $\rho = 0.8$ และ $n = 20$

เมื่อ $K = 3$ $\rho = 0.8$ และ $n = 100$ พบว่าค่าเฉลี่ยของค่า Purity ค่า Rand statistic และค่า Jaccard coefficient จากการวิเคราะห์กลุ่มด้วยของ KR ระยะห่างของ P ระยะห่างแบบ KR&N และระยะห่างแบบ P&N มีค่าจากมากไปน้อยตามลำดับ โดยบางส่วนของช่วงความเชื่อมั่น 95% ทับซ้อนกัน ดังรูปที่ 4.12

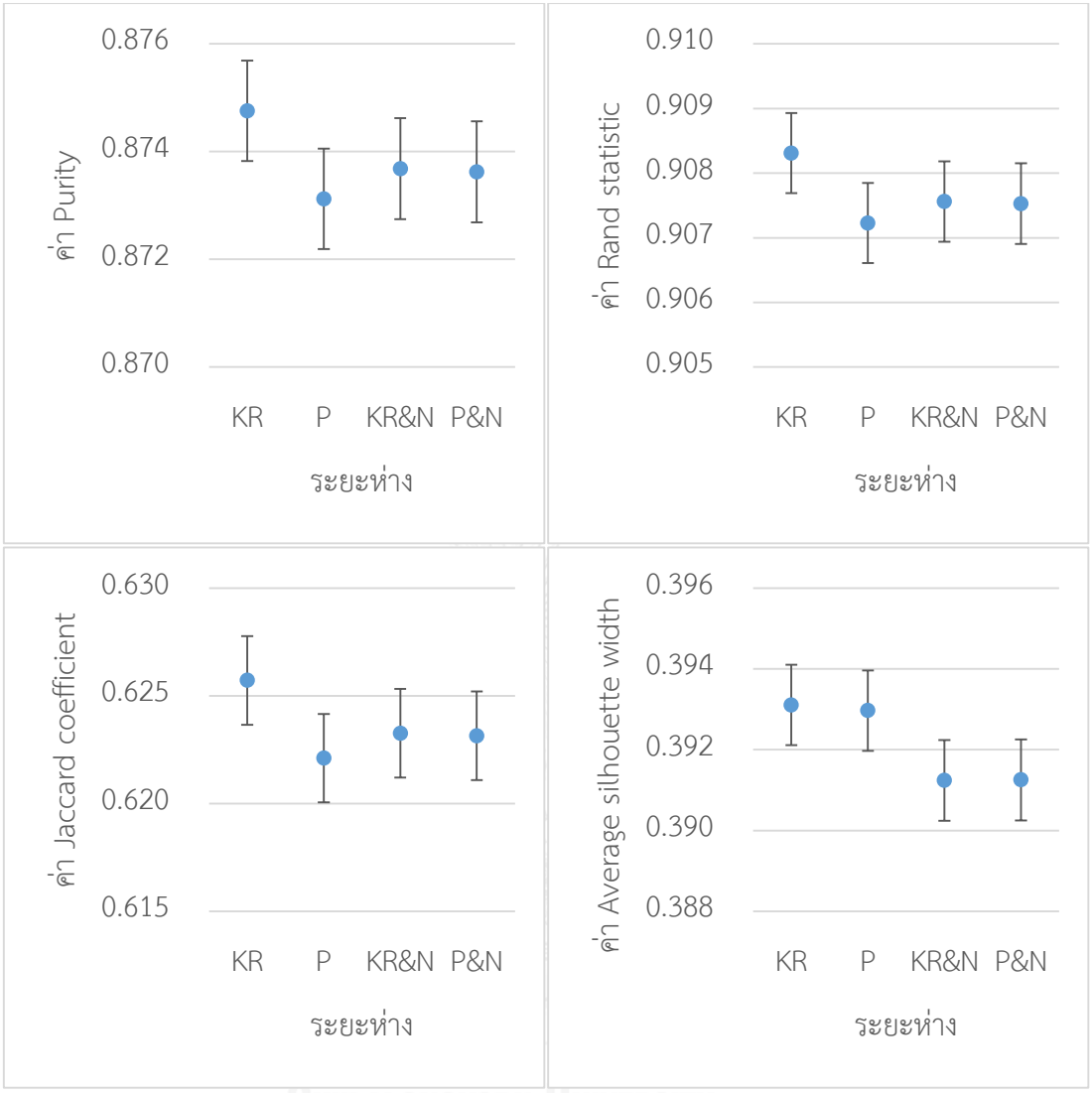


รูปที่ 4.12 กราฟช่วงความเชื่อมั่น 95% ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ II เมื่อ $K = 3$, $\rho = 0.8$ และ $n = 100$

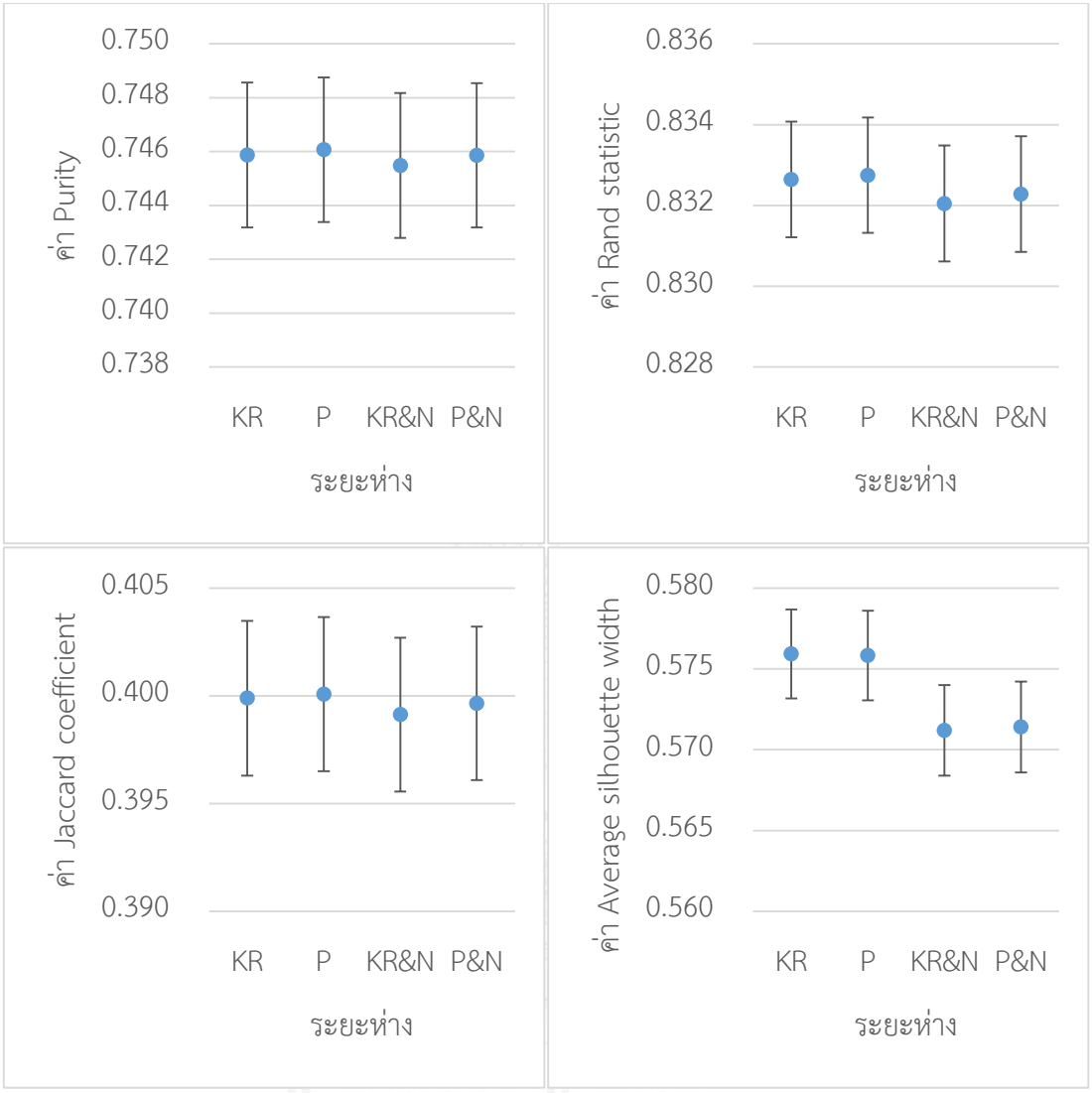
เมื่อ K = 5 แม้ว่าค่าเฉลี่ยของ ค่า Purity ค่า Rand statistic และค่า Jaccard coefficient จากการวิเคราะห์กลุ่มด้วยระยะห่างทั้ง 4 วิธี จะมีค่าแตกต่างกัน แต่เมื่อพิจารณากราฟช่วงความเชื่อมั่น 95% พบว่า การวิเคราะห์กลุ่มด้วยระยะห่างทั้ง 4 วิธี มีประสิทธิภาพโดยเฉลี่ยใกล้เคียงกัน และมีส่วนของช่วงความเชื่อมั่น 95% ทับซ้อนกันในทุกกรณี นอกจากนี้ค่าเฉลี่ยของค่า Average silhouette width จากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR และระยะห่างของ P มีค่าใกล้เคียงกันและมากกว่าจากการวิเคราะห์กลุ่มด้วยระยะห่างแบบ KR&N และระยะห่างแบบ P&N ซึ่งมีค่าใกล้เคียงกันเช่นกัน ดังรูปที่ 4.13 ถึง 4.16



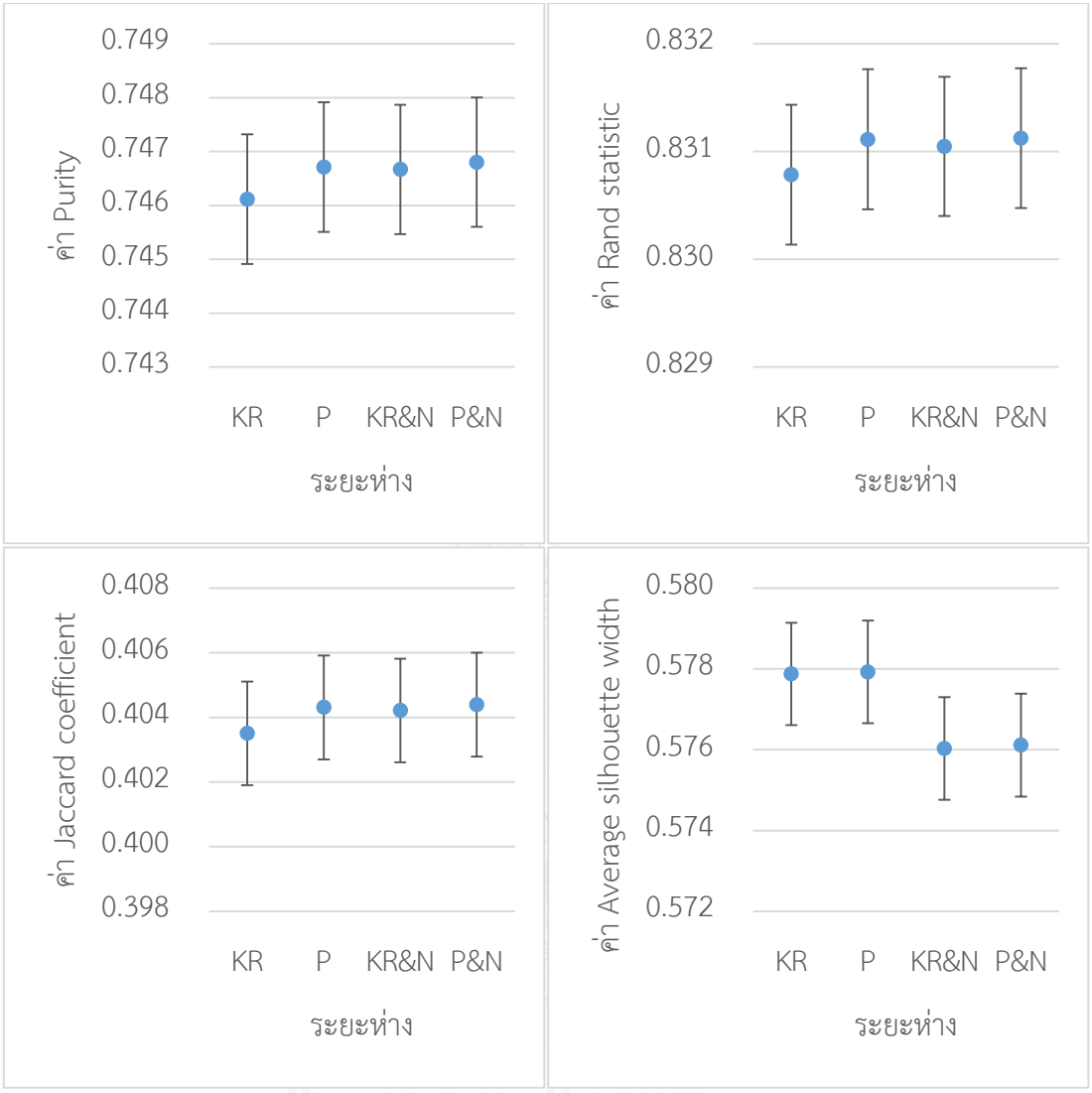
รูปที่ 4.13 กราฟช่วงความเชื่อมั่น 95% ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ II เมื่อ K = 5, ρ = 0.2 และ n = 20



รูปที่ 4.14 กราฟช่วงความเชื่อมั่น 95% ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ II เมื่อ $K = 5$, $\rho = 0.2$ และ $n = 100$



รูปที่ 4.15 กราฟช่วงความเชื่อมั่น 95% ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ II เมื่อ $K = 5$, $\rho = 0.8$ และ $n = 20$



รูปที่ 4.16 กราฟช่วงความเชื่อมั่น 95% ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ II เมื่อ $K = 5$, $\rho = 0.8$ และ $n = 100$

เนื่องจากการพิจารณากราฟช่วงความเชื่อมั่น 95% ของข้อมูลรูปแบบที่ II พบว่ามีบางกรณีที่ค่าเฉลี่ยของค่าวัดประสิทธิภาพการวิเคราะห์กลุ่มใกล้เคียงกัน และมีช่วงความเชื่อมั่น 95% ทับซ้อนกัน ผู้วิจัยจึงพิจารณาความแตกต่างของประสิทธิภาพการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ ในระดับนัยสำคัญ โดยแบ่งเป็นกรณีใหญ่ 2 กรณีตามจำนวนกลุ่มข้อมูล คือกรณีที่ $K = 3$ และ $K = 5$

4.1.2.1 ผลทดสอบความแตกต่างการวิเคราะห์กลุ่มข้อมูลรูปแบบที่ II เมื่อ $K = 3$

ข้อมูลรูปแบบที่ II เมื่อ $K = 3$, $\rho = 0.2$ และ $n = 20$

สำหรับข้อมูลรูปแบบที่ II เมื่อ $K = 3$, $\rho = 0.2$ และ $n = 20$ ทำการวิเคราะห์ความแปรปรวน พบว่าค่า Purity ค่า Rand statistic ค่า Jaccard coefficient และค่า Average silhouette width ต่างมี Sig. = 0.0000 ดังตารางที่ 4.19 ซึ่งน้อยกว่า 0.05 จึงปฏิเสธสมมติฐานหลัก นั่นคือ ประสิทธิภาพการวิเคราะห์กลุ่มขึ้นกับระยะห่างที่ใช้ในการวิเคราะห์กลุ่มที่ระดับนัยสำคัญ 0.05 และค่า Average silhouette width ก็มีค่าขึ้นกับระยะห่างที่ระดับนัยสำคัญ 0.05 เช่นเดียวกัน

ตารางที่ 4.19 ผลวิเคราะห์ความแปรปรวนของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ II เมื่อ $K = 3$, $\rho = 0.2$ และ $n = 20$

แหล่งความแปรปรวน		SS	df	MS	F	Sig.
ค่า Purity	ระหว่างกลุ่ม	0.2548	3	0.0849	61.5218	**0.0000**
	ภายในกลุ่ม	5.5167	3996	0.0014		
	ผลรวม	5.7715	3999			
ค่า Rand statistic	ระหว่างกลุ่ม	0.3652	3	0.1217	64.5958	**0.0000**
	ภายในกลุ่ม	7.5302	3996	0.0019		
	ผลรวม	7.8954	3999			
ค่า Jaccard coefficient	ระหว่างกลุ่ม	1.7422	3	0.5807	62.6970	**0.0000**
	ภายในกลุ่ม	37.0125	3996	0.0093		
	ผลรวม	38.7547	3999			
ค่า Average silhouette width	ระหว่างกลุ่ม	9.2744	3	3.0915	1537.9313	**0.0000**
	ภายในกลุ่ม	8.0326	3996	0.0020		
	ผลรวม	17.3070	3999			

ผู้วิจัยจึงทดสอบความแตกต่างระหว่างค่าเฉลี่ยของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแต่ละคู่ด้วยวิธี Tukey พบว่าค่าเฉลี่ยของค่า Purity ค่า Rand statistic และค่า Jaccard coefficient ที่ได้จากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR และระยะห่างของ P ไม่แตกต่างกันเพียงคู่เดียว แสดงว่าโดยเฉลี่ยการวิเคราะห์กลุ่มด้วยระยะห่างของ KR และระยะห่างของ P มีประสิทธิภาพไม่แตกต่างกัน ขณะที่การวิเคราะห์กลุ่มด้วยระยะห่างวิธีอื่น ๆ มีประสิทธิภาพ

แตกต่างกันที่ระดับนัยสำคัญ 0.05 นอกจากนี้ค่า Average silhouette width จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ ทั้ง 4 วิธีนี้ แตกต่างกันที่ระดับนัยสำคัญ 0.05 ดังตารางที่ 4.20

ตารางที่ 4.20 ผลทดสอบความแตกต่างระหว่างค่าเฉลี่ยแต่ละคู่ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ II เมื่อ $K = 3$, $\rho = 0.2$ และ $n = 20$ ด้วยวิธี Tukey

ค่าที่วัด	ระยะห่าง		ผลต่างค่าเฉลี่ย	Std. Error	Sig.
ค่า Purity	KR	P	-0.0037	0.0017	0.1136
		KR&N	0.0111	0.0017	**0.0000**
		P&N	0.0159	0.0017	**0.0000**
	P	KR	0.0037	0.0017	0.1136
		KR&N	0.0148	0.0017	**0.0000**
		P&N	0.0197	0.0017	**0.0000**
	KR&N	KR	-0.0111	0.0017	**0.0000**
		P	-0.0148	0.0017	**0.0000**
		P&N	0.0049	0.0017	**0.0185**
	P&N	KR	-0.0159	0.0017	**0.0000**
		P	-0.0197	0.0017	**0.0000**
		KR&N	-0.0049	0.0017	**0.0185**
ค่า Rand statistic	KR	P	-0.0043	0.0019	0.1236
		KR&N	0.0134	0.0019	**0.0000**
		P&N	0.0192	0.0019	**0.0000**
	P	KR	0.0043	0.0019	0.1236
		KR&N	0.0176	0.0019	**0.0000**
		P&N	0.0235	0.0019	**0.0000**
	KR&N	KR	-0.0134	0.0019	**0.0000**
		P	-0.0176	0.0019	**0.0000**
		P&N	0.0058	0.0019	**0.0142**
	P&N	KR	-0.0192	0.0019	**0.0000**
		P	-0.0235	0.0019	**0.0000**
		KR&N	-0.0058	0.0019	**0.0142**

ตารางที่ 4.20 (ต่อ) ผลทดสอบความแตกต่างระหว่างค่าเฉลี่ยแต่ละคู่ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ II เมื่อ K = 3, ρ = 0.2 และ n = 20 ด้วยวิธี Tukey

ค่าที่วัด	ระยะห่าง		ผลต่างค่าเฉลี่ย	Std. Error	Sig.
ค่า Jaccard coefficient	KR	P	-0.0102	0.0043	0.0820
		KR&N	0.0286	0.0043	**0.0000**
		P&N	0.0414	0.0043	**0.0000**
	P	KR	0.0102	0.0043	0.0820
		KR&N	0.0389	0.0043	**0.0000**
		P&N	0.0516	0.0043	**0.0000**
	KR&N	KR	-0.0286	0.0043	**0.0000**
		P	-0.0389	0.0043	**0.0000**
		P&N	0.0127	0.0043	**0.0166**
	P&N	KR	-0.0414	0.0043	**0.0000**
		P	-0.0516	0.0043	**0.0000**
		KR&N	-0.0127	0.0043	**0.0166**
ค่า Average silhouette width	KR	P	-0.0079	0.0020	**0.0004**
		KR&N	0.1035	0.0020	**0.0000**
		P&N	0.0772	0.0020	**0.0000**
	P	KR	0.0079	0.0020	**0.0004**
		KR&N	0.1114	0.0020	**0.0000**
		P&N	0.0852	0.0020	**0.0000**
	KR&N	KR	-0.1035	0.0020	**0.0000**
		P	-0.1114	0.0020	**0.0000**
		P&N	-0.0263	0.0020	**0.0000**
	P&N	KR	-0.0772	0.0020	**0.0000**
		P	-0.0852	0.0020	**0.0000**
		KR&N	0.0263	0.0020	**0.0000**

ข้อมูลรูปแบบที่ II เมื่อ $K = 3,\ \rho = 0.2$ *และ* $n = 100$

ผลการวิเคราะห์ความแปรปรวน พบว่าค่า Purity ค่า Rand statistic ค่า Jaccard coefficient และค่า Average silhouette width ต่างมี Sig. = 0.0000 ดังตารางที่ 4.21 ซึ่งน้อยกว่า 0.05 จึงปฏิเสธสมมติฐานหลัก นั่นคือ ประสิทธิภาพการวิเคราะห์กลุ่มขึ้นกับระยะห่างที่ใช้ในการวิเคราะห์กลุ่มที่ระดับนัยสำคัญ 0.05 และค่า Average silhouette width ก็มีค่าขึ้นกับระยะห่างที่ระดับนัยสำคัญ 0.05 เช่นเดียวกัน

ตารางที่ 4.21 ผลวิเคราะห์ความแปรปรวนของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ II เมื่อ $K = 3,\ \rho = 0.2$ และ $n = 100$

แหล่งความแปรปรวน		SS	df	MS	F	Sig.
ค่า Purity	ระหว่างกลุ่ม	0.2073	3	0.0691	289.0555	**0.0000**
	ภายในกลุ่ม	0.9554	3996	0.0002		
	ผลรวม	1.1627	3999			
ค่า Rand statistic	ระหว่างกลุ่ม	0.3082	3	0.1027	311.7191	**0.0000**
	ภายในกลุ่ม	1.3170	3996	0.0003		
	ผลรวม	1.6252	3999			
ค่า Jaccard coefficient	ระหว่างกลุ่ม	1.4177	3	0.4726	299.3697	**0.0000**
	ภายในกลุ่ม	6.3079	3996	0.0016		
	ผลรวม	7.7256	3999			
ค่า Average silhouette width	ระหว่างกลุ่ม	11.8365	3	3.9455	10855.3551	**0.0000**
	ภายในกลุ่ม	1.4524	3996	0.0004		
	ผลรวม	13.2889	3999			

ผู้วิจัยทดสอบความแตกต่างของค่าเฉลี่ยของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแต่ละคู่ ด้วยวิธี Tukey พบว่าค่าเฉลี่ยของค่า Purity ค่า Rand statistic ค่า Jaccard coefficient และค่า Average silhouette width ที่ได้จากการวิเคราะห์กลุ่มด้วยระยะห่างทั้ง 4 วิธี แตกต่างกันที่ระดับนัยสำคัญ 0.05 นั่นคือโดยเฉลี่ยการวิเคราะห์กลุ่มด้วยระยะห่างทั้ง 4 วิธีมีประสิทธิภาพแตกต่างกันอย่างมีนัยสำคัญ และค่า Average silhouette width จากการวิเคราะห์กลุ่มด้วยระยะห่าง ทั้ง 4 วิธีนี้ แตกต่างกันอย่างมีนัยสำคัญ ดังตารางที่ 4.22

ตารางที่ 4.22 ผลทดสอบความแตกต่างระหว่างค่าเฉลี่ยแต่ละคู่ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ II เมื่อ K = 3, ρ = 0.2 และ n = 100 ด้วยวิธี Tukey

ค่าที่วัด	ระยะห่าง		ผลต่างค่าเฉลี่ย	Std. Error	Sig.
ค่า Purity	KR	P	-0.0036	0.0007	**0.0000**
		KR&N	0.0063	0.0007	**0.0000**
		P&N	0.0154	0.0007	**0.0000**
	P	KR	0.0036	0.0007	**0.0000**
		KR&N	0.0099	0.0007	**0.0000**
		P&N	0.0190	0.0007	**0.0000**
	KR&N	KR	-0.0063	0.0007	**0.0000**
		P	-0.0099	0.0007	**0.0000**
		P&N	0.0091	0.0007	**0.0000**
	P&N	KR	-0.0154	0.0007	**0.0000**
		P	-0.0190	0.0007	**0.0000**
		KR&N	-0.0091	0.0007	**0.0000**
ค่า Rand statistic	KR	P	-0.0040	0.0008	**0.0000**
		KR&N	0.0081	0.0008	**0.0000**
		P&N	0.0189	0.0008	**0.0000**
	P	KR	0.0040	0.0008	**0.0000**
		KR&N	0.0121	0.0008	**0.0000**
		P&N	0.0230	0.0008	**0.0000**
	KR&N	KR	-0.0081	0.0008	**0.0000**
		P	-0.0121	0.0008	**0.0000**
		P&N	0.0109	0.0008	**0.0000**
	P&N	KR	-0.0189	0.0008	**0.0000**
		P	-0.0230	0.0008	**0.0000**
		KR&N	-0.0109	0.0008	**0.0000**

ตารางที่ 4.22 (ต่อ) ผลทดสอบความแตกต่างระหว่างค่าเฉลี่ยแต่ละคู่ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ II เมื่อ K = 3, ρ = 0.2 และ n = 100 ด้วยวิธี Tukey

ค่าที่วัด	ระยะห่าง		ผลต่างค่าเฉลี่ย	Std. Error	Sig.
ค่า Jaccard coefficient	KR	P	-0.0094	0.0018	**0.0000**
		KR&N	0.0170	0.0018	**0.0000**
		P&N	0.0401	0.0018	**0.0000**
	P	KR	0.0094	0.0018	**0.0000**
		KR&N	0.0265	0.0018	**0.0000**
		P&N	0.0495	0.0018	**0.0000**
	KR&N	KR	-0.0170	0.0018	**0.0000**
		P	-0.0265	0.0018	**0.0000**
		P&N	0.0231	0.0018	**0.0000**
	P&N	KR	-0.0401	0.0018	**0.0000**
		P	-0.0495	0.0018	**0.0000**
		KR&N	-0.0231	0.0018	**0.0000**
ค่า Average silhouette width	KR	P	-0.0078	0.0009	**0.0000**
		KR&N	0.1168	0.0009	**0.0000**
		P&N	0.0892	0.0009	**0.0000**
	P	KR	0.0078	0.0009	**0.0000**
		KR&N	0.1246	0.0009	**0.0000**
		P&N	0.0970	0.0009	**0.0000**
	KR&N	KR	-0.1168	0.0009	**0.0000**
		P	-0.1246	0.0009	**0.0000**
		P&N	-0.0276	0.0009	**0.0000**
	P&N	KR	-0.0892	0.0009	**0.0000**
		P	-0.0970	0.0009	**0.0000**
		KR&N	0.0276	0.0009	**0.0000**

ข้อมูลรูปแบบที่ II เมื่อ $K = 3,\ \rho = 0.8$ *และ* $n = 20$

ผลการวิเคราะห์ความแปรปรวน พบว่าค่า Purity ค่า Rand statistic ค่า Jaccard coefficient และค่า Average silhouette width ต่างมี Sig. = 0.0000 ดังตารางที่ 4.23 ซึ่งน้อยกว่า 0.05 จึงปฏิเสธสมมติฐานหลัก นั่นคือ ประสิทธิภาพการวิเคราะห์กลุ่มและค่า Average silhouette width ขึ้นกับระยะห่างที่ใช้ในการวิเคราะห์กลุ่มที่ระดับนัยสำคัญ 0.05

ตารางที่ 4.23 ผลวิเคราะห์ความแปรปรวนของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ II เมื่อ $K = 3,\ \rho = 0.8$ และ $n = 20$

แหล่งความแปรปรวน		SS	df	MS	F	Sig.
ค่า Purity	ระหว่างกลุ่ม	0.0911	3	0.0304	12.2139	**0.0000**
	ภายในกลุ่ม	9.9389	3996	0.0025		
	ผลรวม	10.0300	3999			
ค่า Rand statistic	ระหว่างกลุ่ม	0.1618	3	0.0539	20.0951	**0.0000**
	ภายในกลุ่ม	10.7226	3996	0.0027		
	ผลรวม	10.8844	3999			
ค่า Jaccard coefficient	ระหว่างกลุ่ม	0.3147	3	0.1049	15.7342	**0.0000**
	ภายในกลุ่ม	26.6401	3996	0.0067		
	ผลรวม	26.9548	3999			
ค่า Average silhouette width	ระหว่างกลุ่ม	8.2313	3	2.7438	904.5848	**0.0000**
	ภายในกลุ่ม	12.1206	3996	0.0030		
	ผลรวม	20.3519	3999			

ผลทดสอบความแตกต่างของค่าเฉลี่ยของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแต่ละคู่ พบว่าค่าเฉลี่ยของค่า Purity ค่า Rand statistic และค่า Jaccard coefficient ที่ได้จากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR และระยะห่างของ P ไม่แตกต่างกัน แต่แตกต่างจากระยะห่างแบบ KR&N และระยะห่างแบบ P&N ที่ระดับนัยสำคัญ 0.05 นอกจากนี้ระยะห่างแบบ KR&N และระยะห่างแบบ P&N ไม่แตกต่างกันอีกด้วย ดังนั้น แม้ว่าการวิเคราะห์กลุ่มด้วยระยะห่างของ P มีประสิทธิภาพดีที่สุดโดยเฉลี่ย แต่การวิเคราะห์กลุ่มด้วยระยะห่างของ P และระยะห่างของ KR มีประสิทธิภาพไม่แตกต่างกัน แต่แตกต่างจากระยะห่างแบบ KR&N และระยะห่างแบบ P&N อย่างมีนัยสำคัญ ขณะที่การวิเคราะห์กลุ่มด้วยระยะห่างแบบ KR&N และระยะห่างแบบ P&N มีประสิทธิภาพไม่แตกต่างกัน นอกจากนี้ค่า

Average silhouette width จากการวิเคราะห์กลุ่มด้วยระยะห่างทั้ง 4 วิธีนี้ แตกต่างกันที่ระดับนัยสำคัญ 0.05 ดังตารางที่ 4.24

ตารางที่ 4.24 ผลทดสอบความแตกต่างระหว่างค่าเฉลี่ยแต่ละคู่ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ II เมื่อ K = 3, ρ = 0.8 และ n = 20 ด้วยวิธี Tukey

ค่าที่วัด	ระยะห่าง		ผลต่างค่าเฉลี่ย	Std. Error	Sig.
ค่า Purity	KR	P	0.0012	0.0022	0.9458
		KR&N	0.0099	0.0022	**0.0001**
		P&N	0.0104	0.0022	**0.0000**
	P	KR	-0.0012	0.0022	0.9458
		KR&N	0.0086	0.0022	**0.0006**
		P&N	0.0091	0.0022	**0.0003**
	KR&N	KR	-0.0099	0.0022	**0.0001**
		P	-0.0086	0.0022	**0.0006**
		P&N	0.0005	0.0022	0.9960
	P&N	KR	-0.0104	0.0022	**0.0000**
		P	-0.0091	0.0022	**0.0003**
		KR&N	-0.0005	0.0022	0.9960
ค่า Rand statistic	KR	P	0.0019	0.0023	0.8373
		KR&N	0.0150	0.0023	**0.0000**
		P&N	0.0119	0.0023	**0.0000**
	P	KR	-0.0019	0.0023	0.8373
		KR&N	0.0130	0.0023	**0.0000**
		P&N	0.0100	0.0023	**0.0001**
	KR&N	KR	-0.0150	0.0023	**0.0000**
		P	-0.0130	0.0023	**0.0000**
		P&N	-0.0031	0.0023	0.5456
	P&N	KR	-0.0119	0.0023	**0.0000**
		P	-0.0100	0.0023	**0.0001**
		KR&N	0.0031	0.0023	0.5456

ตารางที่ 4.24 (ต่อ) ผลทดสอบความแตกต่างระหว่างค่าเฉลี่ยแต่ละคู่ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ II เมื่อ K = 3, ρ = 0.8 และ n = 20 ด้วยวิธี Tukey

ค่าที่วัด	ระยะห่าง		ผลต่างค่าเฉลี่ย	Std. Error	Sig.
ค่า Jaccard coefficient	KR	P	0.0028	0.0037	0.8708
		KR&N	0.0199	0.0037	**0.0000**
		P&N	0.0180	0.0037	**0.0000**
	P	KR	-0.0028	0.0037	0.8708
		KR&N	0.0171	0.0037	**0.0000**
		P&N	0.0153	0.0037	**0.0002**
	KR&N	KR	-0.0199	0.0037	**0.0000**
		P	-0.0171	0.0037	**0.0000**
		P&N	-0.0019	0.0037	0.9564
	P&N	KR	-0.0180	0.0037	**0.0000**
		P	-0.0153	0.0037	**0.0002**
		KR&N	0.0019	0.0037	0.9564
ค่า Average silhouette width	KR	P	-0.0156	0.0025	**0.0000**
		KR&N	0.0971	0.0025	**0.0000**
		P&N	0.0593	0.0025	**0.0000**
	P	KR	0.0156	0.0025	**0.0000**
		KR&N	0.1127	0.0025	**0.0000**
		P&N	0.0749	0.0025	**0.0000**
	KR&N	KR	-0.0971	0.0025	**0.0000**
		P	-0.1127	0.0025	**0.0000**
		P&N	-0.0378	0.0025	**0.0000**
	P&N	KR	-0.0593	0.0025	**0.0000**
		P	-0.0749	0.0025	**0.0000**
		KR&N	0.0378	0.0025	**0.0000**

ข้อมูลรูปแบบที่ II เมื่อ $K = 3,\ \rho = 0.8$ *และ* $n = 100$

ผลการวิเคราะห์ความแปรปรวน พบว่าค่า Purity ค่า Rand statistic ค่า Jaccard coefficient และค่า Average silhouette width ต่างมี Sig. = 0.0000 ดังตารางที่ 4.25 ซึ่งน้อยกว่า 0.05 จึงปฏิเสธสมมติฐานหลัก นั่นคือประสิทธิภาพการวิเคราะห์กลุ่มขึ้นกับระยะห่างที่ใช้ในการวิเคราะห์กลุ่มที่ระดับนัยสำคัญ 0.05 และค่า Average silhouette width ก็มีค่าขึ้นกับระยะห่างที่ระดับนัยสำคัญ 0.05 เช่นเดียวกัน

ตารางที่ 4.25 ผลวิเคราะห์ความแปรปรวนของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ II เมื่อ $K = 3,\ \rho = 0.8$ และ $n = 100$

แหล่งความแปรปรวน		SS	df	MS	F	Sig.
ค่า Purity	ระหว่างกลุ่ม	0.0599	3	0.0200	39.0356	**0.0000**
	ภายในกลุ่ม	2.0443	3996	0.0005		
	ผลรวม	2.1042	3999			
ค่า Rand statistic	ระหว่างกลุ่ม	0.0728	3	0.0243	47.2757	**0.0000**
	ภายในกลุ่ม	2.0517	3996	0.0005		
	ผลรวม	2.1245	3999			
ค่า Jaccard coefficient	ระหว่างกลุ่ม	0.1530	3	0.0510	40.5586	**0.0000**
	ภายในกลุ่ม	5.0255	3996	0.0013		
	ผลรวม	5.1785	3999			
ค่า Average silhouette width	ระหว่างกลุ่ม	13.2280	3	4.4093	7141.7221	**0.0000**
	ภายในกลุ่ม	2.4672	3996	0.0006		
	ผลรวม	15.6952	3999			

ผลทดสอบความแตกต่างของค่าเฉลี่ยของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแต่ละคู่ พบว่าโดยเฉลี่ย ค่า Purity และค่า Jaccard coefficient ที่ได้จากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR และระยะห่างของ P ไม่แตกต่างกัน แต่แตกต่างจากระยะห่างแบบ KR&N และระยะห่างแบบ P&N ที่ระดับนัยสำคัญ 0.05 ขณะที่ค่า Rand statistic ที่ได้จากการวิเคราะห์กลุ่มด้วยระยะห่างทั้ง 4 วิธีนี้ต่างก็แตกต่างกันที่ระดับนัยสำคัญ 0.05 นอกจากนี้ค่า Average silhouette width จากการวิเคราะห์กลุ่มด้วยระยะห่างทั้ง 4 วิธีนี้ แตกต่างกันที่ระดับนัยสำคัญ 0.05 ดังตารางที่ 4.26

ตารางที่ 4.26 ผลทดสอบความแตกต่างระหว่างค่าเฉลี่ยแต่ละคู่ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ II เมื่อ K = 3, ρ = 0.8 และ n = 100 ด้วยวิธี Tukey

ค่าที่วัด	ระยะห่าง		ผลต่างค่าเฉลี่ย	Std. Error	Sig.
ค่า Purity	KR	P	0.0021	0.0010	0.1611
		KR&N	0.0066	0.0010	**0.0000**
		P&N	0.0099	0.0010	**0.0000**
	P	KR	-0.0021	0.0010	0.1611
		KR&N	0.0045	0.0010	**0.0001**
		P&N	0.0078	0.0010	**0.0000**
	KR&N	KR	-0.0066	0.0010	**0.0000**
		P	-0.0045	0.0010	**0.0001**
		P&N	0.0033	0.0010	**0.0054**
	P&N	KR	-0.0099	0.0010	**0.0000**
		P	-0.0078	0.0010	**0.0000**
		KR&N	-0.0033	0.0010	**0.0054**
ค่า Rand statistic	KR	P	0.0027	0.0010	**0.0393**
		KR&N	0.0077	0.0010	**0.0000**
		P&N	0.0110	0.0010	**0.0000**
	P	KR	-0.0027	0.0010	**0.0393**
		KR&N	0.0050	0.0010	**0.0000**
		P&N	0.0083	0.0010	**0.0000**
	KR&N	KR	-0.0077	0.0010	**0.0000**
		P	-0.0050	0.0010	**0.0000**
		P&N	0.0033	0.0010	**0.0068**
	P&N	KR	-0.0110	0.0010	**0.0000**
		P	-0.0083	0.0010	**0.0000**
		KR&N	-0.0033	0.0010	**0.0068**

ตารางที่ 4.26 (ต่อ) ผลทดสอบความแตกต่างระหว่างค่าเฉลี่ยแต่ละคู่ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ II เมื่อ K = 3, ρ = 0.8 และ n = 100 ด้วยวิธี Tukey

ค่าที่วัด	ระยะห่าง		ผลต่างค่าเฉลี่ย	Std. Error	Sig.
ค่า Jaccard coefficient	KR	P	0.0034	0.0016	0.1307
		KR&N	0.0108	0.0016	**0.0000**
		P&N	0.0158	0.0016	**0.0000**
	P	KR	-0.0034	0.0016	0.1307
		KR&N	0.0073	0.0016	**0.0000**
		P&N	0.0124	0.0016	**0.0000**
	KR&N	KR	-0.0108	0.0016	**0.0000**
		P	-0.0073	0.0016	**0.0000**
		P&N	0.0051	0.0016	**0.0074**
	P&N	KR	-0.0158	0.0016	**0.0000**
		P	-0.0124	0.0016	**0.0000**
		KR&N	-0.0051	0.0016	**0.0074**
ค่า Average silhouette width	KR	P	-0.0172	0.0011	**0.0000**
		KR&N	0.1242	0.0011	**0.0000**
		P&N	0.0778	0.0011	**0.0000**
	P	KR	0.0172	0.0011	**0.0000**
		KR&N	0.1414	0.0011	**0.0000**
		P&N	0.0950	0.0011	**0.0000**
	KR&N	KR	-0.1242	0.0011	**0.0000**
		P	-0.1414	0.0011	**0.0000**
		P&N	-0.0464	0.0011	**0.0000**
	P&N	KR	-0.0778	0.0011	**0.0000**
		P	-0.0950	0.0011	**0.0000**
		KR&N	0.0464	0.0011	**0.0000**

4.1.2.2 ผลทดสอบความแตกต่างการวิเคราะห์กลุ่มข้อมูลรูปแบบที่ II เมื่อ $K = 5$

เนื่องจากเมื่อ $K = 5$ ยังสามารถแบ่งข้อมูลได้อีกตามความแตกต่างของค่าสัมประสิทธิ์สหสัมพันธ์ และขนาดข้อมูลต่อกลุ่ม จึงทดสอบความแตกต่างประสิทธิภาพการวิเคราะห์กลุ่มข้อมูลต่าง ๆ ได้ดังต่อไปนี้

ข้อมูลรูปแบบที่ II เมื่อ $K = 5$, $\rho = 0.2$ และ $n = 20$

ผลวิเคราะห์ความแปรปรวน ดังตารางที่ 4.27 พบว่าค่า Purity ค่า Rand statistic และค่า Jaccard coefficient มี Sig. มากกว่า 0.05 จึงไม่สามารถปฏิเสธสมมติฐานหลัก ดังนั้นประสิทธิภาพการวิเคราะห์กลุ่มไม่ขึ้นกับระยะห่างที่ใช้ในการวิเคราะห์กลุ่ม อย่างไรก็ตามค่า Average silhouette width มี Sig. = 0.0000 ซึ่งน้อยกว่า 0.05 จึงปฏิเสธสมมติฐานหลัก แสดงว่าค่า Average silhouette width มีค่าขึ้นกับระยะห่างที่ระดับนัยสำคัญ 0.05

ตารางที่ 4.27 ผลวิเคราะห์ความแปรปรวนของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ II เมื่อ $K = 5$, $\rho = 0.2$ และ $n = 20$

แหล่งความแปรปรวน		SS	df	MS	F	Sig.
ค่า Purity	ระหว่างกลุ่ม	0.0068	3	0.0023	1.9785	0.1150
	ภายในกลุ่ม	4.5722	3996	0.0011		
	ผลรวม	4.5790	3999			
ค่า Rand statistic	ระหว่างกลุ่ม	0.0036	3	0.0012	2.4274	0.0636
	ภายในกลุ่ม	1.9741	3996	0.0005		
	ผลรวม	1.9777	3999			
ค่า Jaccard coefficient	ระหว่างกลุ่ม	0.0330	3	0.0110	1.9938	0.1127
	ภายในกลุ่ม	22.0266	3996	0.0055		
	ผลรวม	22.0595	3999			
ค่า Average silhouette width	ระหว่างกลุ่ม	0.0367	3	0.0122	8.8145	**0.0000**
	ภายในกลุ่ม	5.5460	3996	0.0014		
	ผลรวม	5.5827	3999			

ผู้วิจัยทดสอบความแตกต่างระหว่างค่าเฉลี่ยแต่ละคู่ ของค่า Average silhouette width จากการวิเคราะห์กลุ่มด้วยระยะห่างต่าง ๆ ด้วยวิธี Tukey ดังตารางที่ 4.28 พบว่า ค่า Average silhouette width จากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR แตกต่างจากระยะห่างแบบ KR&N และระยะห่างแบบ P&N ที่ระดับนัยสำคัญ 0.05 แต่ไม่แตกต่างจากระยะห่างของ P ส่วนค่า Average silhouette width จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบ KR&N แตกต่างจากระยะห่างของ KR และระยะห่างของ P ที่ระดับนัยสำคัญ 0.05 แต่ไม่แตกต่างจากระยะห่างแบบ P&N

ตารางที่ 4.28 ผลทดสอบความแตกต่างระหว่างค่าเฉลี่ยแต่ละคู่ของค่า Average silhouette width จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ II เมื่อ K = 5, ρ = 0.2 และ n = 20 ด้วยวิธี Tukey

ระยะห่าง		ผลต่างค่าเฉลี่ย	Std. Error	Sig.
KR	P	0.0000	0.0017	1.0000
	KR&N	0.0061	0.0017	**0.0014**
	P&N	0.0060	0.0017	**0.0019**
P	KR	0.0000	0.0017	1.0000
	KR&N	0.0061	0.0017	**0.0013**
	P&N	0.0060	0.0017	**0.0018**
KR&N	KR	-0.0061	0.0017	**0.0014**
	P	-0.0061	0.0017	**0.0013**
	P&N	-0.0001	0.0017	0.9998
P&N	KR	-0.0060	0.0017	**0.0019**
	P	-0.0060	0.0017	**0.0018**
	KR&N	0.0001	0.0017	0.9998

ข้อมูลรูปแบบที่ II เมื่อ $K = 5,\ \rho = 0.2$ *และ* $n = 100$

ผลวิเคราะห์ความแปรปรวน ดังตารางที่ 4.29 พบว่าค่า Purity ค่า Rand statistic และค่า Jaccard coefficient มี Sig. มากกว่า 0.05 จึงไม่สามารถปฏิเสธสมมติฐานหลัก ดังนั้น ประสิทธิภาพการวิเคราะห์กลุ่มไม่ขึ้นกับระยะห่างที่ใช้ในการวิเคราะห์กลุ่ม อย่างไรก็ตามค่า Average silhouette width มี Sig. น้อยกว่า 0.05 จึงปฏิเสธสมมติฐานหลัก ดังนั้น ค่า Average silhouette width มีค่าขึ้นกับระยะห่างที่ใช้ในการวิเคราะห์กลุ่มที่ระดับนัยสำคัญ 0.05

ตารางที่ 4.29 ผลวิเคราะห์ความแปรปรวนของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ II เมื่อ $K = 5,\ \rho = 0.2$ และ $n = 100$

แหล่งความแปรปรวน		SS	df	MS	F	Sig.
ค่า Purity	ระหว่างกลุ่ม	0.0014	3	0.0005	2.0814	0.1005
	ภายในกลุ่ม	0.9080	3996	0.0002		
	ผลรวม	0.9094	3999			
ค่า Rand statistic	ระหว่างกลุ่ม	0.0006	3	0.0002	2.1130	0.0964
	ภายในกลุ่ม	0.4013	3996	0.0001		
	ผลรวม	0.4019	3999			
ค่า Jaccard coefficient	ระหว่างกลุ่ม	0.0070	3	0.0023	2.1285	0.0945
	ภายในกลุ่ม	4.3912	3996	0.0011		
	ผลรวม	4.3982	3999			
ค่า Average silhouette width	ระหว่างกลุ่ม	0.0032	3	0.0011	4.1469	**0.0061**
	ภายในกลุ่ม	1.0335	3996	0.0003		
	ผลรวม	1.0367	3999			

ผู้วิจัยทดสอบความแตกต่างระหว่างค่าเฉลี่ยแต่ละคู่ของค่า Average silhouette width จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ ด้วยวิธี Tukey ดังตารางที่ 4.30 พบว่า ค่า Average silhouette width จากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR แตกต่างจากระยะห่างแบบ KR&N และระยะห่างแบบ P&N ที่ระดับนัยสำคัญ 0.05 แต่ไม่แตกต่างจากระยะห่างของ P ส่วนค่า Average silhouette width จากการวิเคราะห์กลุ่มด้วยระยะห่างของ P ระยะห่างแบบ KR&N และระยะห่างแบบ P&N ไม่แตกต่างกัน

ตารางที่ 4.30 ผลทดสอบความแตกต่างระหว่างค่าเฉลี่ยแต่ละคู่ของค่า Average silhouette width จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ II เมื่อ K = 5, ρ = 0.2 และ n = 100 ด้วยวิธี Tukey

ระยะห่าง		ผลต่างค่าเฉลี่ย	Std. Error	Sig.
KR	P	0.0001	0.0007	0.9974
	KR&N	0.0019	0.0007	**0.0468**
	P&N	0.0019	0.0007	**0.0487**
P	KR	-0.0001	0.0007	0.9974
	KR&N	0.0017	0.0007	0.0772
	P&N	0.0017	0.0007	0.0800
KR&N	KR	-0.0019	0.0007	**0.0468**
	P	-0.0017	0.0007	0.0772
	P&N	0.0000	0.0007	1.0000
P&N	KR	-0.0019	0.0007	**0.0487**
	P	-0.0017	0.0007	0.0800
	KR&N	0.0000	0.0007	1.0000

ข้อมูลรูปแบบที่ II เมื่อ $K = 5,\ \rho = 0.8$ *และ* $n = 20$

ผลวิเคราะห์ความแปรปรวน ดังตารางที่ 4.31 พบว่าค่า Purity ค่า Rand statistic และค่า Jaccard coefficient มี Sig. มากกว่า 0.05 จึงไม่สามารถปฏิเสธสมมติฐานหลัก ดังนั้น ประสิทธิภาพการวิเคราะห์กลุ่มไม่ขึ้นกับระยะห่างที่ใช้ในการวิเคราะห์กลุ่ม อย่างไรก็ตามค่า Average silhouette width มี Sig. = 0.0000 ซึ่งน้อยกว่า 0.05 จึงปฏิเสธสมมติฐานหลัก แสดงว่าค่า Average silhouette width มีค่าขึ้นกับระยะห่างที่ระดับนัยสำคัญ 0.05

ตารางที่ 4. 31 ผลวิเคราะห์ความแปรปรวนของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ II เมื่อ $K = 5,\ \rho = 0.8$ และ $n = 20$

แหล่งความแปรปรวน		SS	df	MS	F	Sig.
ค่า Purity	ระหว่างกลุ่ม	0.0002	3	0.0001	0.0312	0.9926
	ภายในกลุ่ม	7.4895	3996	0.0019		
	ผลรวม	7.4897	3999			
ค่า Rand statistic	ระหว่างกลุ่ม	0.0003	3	0.0001	0.1970	0.8985
	ภายในกลุ่ม	2.1259	3996	0.0005		
	ผลรวม	2.1262	3999			
ค่า Jaccard coefficient	ระหว่างกลุ่ม	0.0005	3	0.0002	0.0509	0.9848
	ภายในกลุ่ม	13.2743	3996	0.0033		
	ผลรวม	13.2748	3999			
ค่า Average silhouette width	ระหว่างกลุ่ม	0.0210	3	0.0070	3.4788	**0.0153**
	ภายในกลุ่ม	8.0446	3996	0.0020		
	ผลรวม	8.0656	3999			

จากผลวิเคราะห์ความแปรปรวน ผู้วิจัยจึงทดสอบความแตกต่างของค่าเฉลี่ยแต่ละคู่ของค่า Average silhouette width จากการวิเคราะห์กลุ่มด้วยระยะห่างต่าง ๆ ด้วยวิธี Tukey พบว่า ความแตกต่างของค่า Average silhouette width จากการวิเคราะห์กลุ่มระหว่างระยะห่างแต่ละคู่จากทั้ง 4 วิธีนี้ มี .Sig มากกว่าหรือเท่ากับ 0.05 ดังตารางที่ 4.32 จึงไม่สามารถปฏิเสธสมมติฐานหลักได้ แสดงว่าค่า Average silhouette width จากการวิเคราะห์กลุ่มระหว่างระยะห่างแต่ละคู่จากทั้ง 4 วิธีนี้ ไม่แตกต่างกัน ซึ่งขัดแย้งกับผลวิเคราะห์ความแปรปรวนก่อนหน้านี้ จึงพิจารณารูปที่ 4.15 เพิ่มเติม พบว่าค่า Average silhouette width จากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR และ

ระยะห่างของ P มีค่าใกล้เคียงกัน โดยมีส่วนที่ทับซ้อนกันของช่วงความเชื่อมั่น 95% และมากกว่าจากการวิเคราะห์กลุ่มด้วยระยะห่างแบบ KR&N และระยะห่างแบบ P&N ซึ่งมีค่าใกล้เคียงกันและมีส่วนที่ทับซ้อนกันของช่วงความเชื่อมั่น 95% เช่นเดียวกัน

ตารางที่ 4.32 ผลทดสอบความแตกต่างระหว่างค่าเฉลี่ยแต่ละคู่ของค่า Average silhouette width จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ II เมื่อ K = 5, ρ = 0.8 และ n = 20 ด้วยวิธี Tukey

ระยะห่าง		ผลต่างค่าเฉลี่ย	Std. Error	Sig.
KR	P	0.0001	0.0020	1.0000
	KR&N	0.0047	0.0020	0.0853
	P&N	0.0045	0.0020	0.1084
P	KR	-0.0001	0.0020	1.0000
	KR&N	0.0046	0.0020	0.0962
	P&N	0.0044	0.0020	0.1215
KR&N	KR	-0.0047	0.0020	0.0853
	P	-0.0046	0.0020	0.0962
	P&N	-0.0002	0.0020	0.9996
P&N	KR	-0.0045	0.0020	0.1084
	P	-0.0044	0.0020	0.1215
	KR&N	0.0002	0.0020	0.9996

ข้อมูลรูปแบบที่ II เมื่อ $K = 5,\ \rho = 0.8$ และ $n = 100$

จากตารางที่ 4.33 และ 4.34 พบว่าสามารถสรุปผลการวิเคราะห์กลุ่มข้อมูลได้เช่นเดียวกับกรณีข้อมูลรูปแบบที่ II เมื่อ $K = 5,\ \rho = 0.8$ และ $n = 20$

ตารางที่ 4.33 ผลวิเคราะห์ความแปรปรวนของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ II เมื่อ $K = 5,\ \rho = 0.8$ และ $n = 100$

แหล่งความแปรปรวน		SS	df	MS	F	Sig.
ค่า Purity	ระหว่างกลุ่ม	0.0003	3	0.0001	0.2573	0.8562
	ภายในกลุ่ม	1.4995	3996	0.0004		
	ผลรวม	1.4998	3999			
ค่า Rand statistic	ระหว่างกลุ่ม	0.0001	3	0.0000	0.2307	0.8750
	ภายในกลุ่ม	0.4362	3996	0.0001		
	ผลรวม	0.4363	3999			
ค่า Jaccard coefficient	ระหว่างกลุ่ม	0.0005	3	0.0002	0.2471	0.8634
	ภายในกลุ่ม	2.6748	3996	0.0007		
	ผลรวม	2.6752	3999			
ค่า Average silhouette width	ระหว่างกลุ่ม	0.0033	3	0.0011	2.6582	**0.0467**
	ภายในกลุ่ม	1.6783	3996	0.0004		
	ผลรวม	1.6816	3999			

ตารางที่ 4.34 ผลทดสอบความแตกต่างระหว่างค่าเฉลี่ยแต่ละคู่ของค่า Average silhouette width จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ II เมื่อ K = 5, ρ = 0.8 และ n = 100 ด้วยวิธี Tukey

ระยะห่าง		ผลต่างค่าเฉลี่ย	Std. Error	Sig.
KR	P	0.0000	0.0009	0.9999
	KR&N	0.0018	0.0009	0.1828
	P&N	0.0018	0.0009	0.2178
P	KR	0.0000	0.0009	0.9999
	KR&N	0.0019	0.0009	0.1644
	P&N	0.0018	0.0009	0.1970
KR&N	KR	-0.0018	0.0009	0.1828
	P	-0.0019	0.0009	0.1644
	P&N	-0.0001	0.0009	0.9997
P&N	KR	-0.0018	0.0009	0.2178
	P	-0.0018	0.0009	0.1970
	KR&N	0.0001	0.0009	0.9997

4.1.3 ผลการวิเคราะห์กลุ่มข้อมูลรูปแบบที่ III

ผู้วิจัยทำการวิเคราะห์กลุ่มด้วยระยะห่าง 2 วิธี ได้แก่ ระยะห่างของ KR และระยะห่างแบบ KR&N สำหรับข้อมูลรูปแบบที่ III เนื่องจากเป็นข้อมูลที่ประกอบไปด้วยตัวแปรเพียง 2 ชนิด คือ ตัวแปรนามบัญญัติ และตัวแปรเชิงปริมาณ อย่างละ 3 ตัวแปร และคำนวณค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของ ค่า Purity ค่า Rand statistic ค่า Jaccard coefficient และค่า Average silhouette width ได้ผลดังตารางที่ 4.34

ผลการเปรียบเทียบประสิทธิภาพโดยเฉลี่ยในการวิเคราะห์กลุ่มด้วยมาตรวัดระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ III พบว่า การวิเคราะห์กลุ่มด้วยระยะห่างแบบ KR&N ให้ค่าเฉลี่ยของค่า Purity ค่า Rand statistic และค่า Jaccard coefficient สูงที่สุด ทุกกรณี ยกเว้นกรณีที่ K = 5 ρ = 0.2 และ n = 20 ดังตารางที่ 4.34

นอกจากนี้ค่าเฉลี่ยของค่า Average silhouette width จากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR มีค่ามากกว่า จากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR&N ในทุกกรณี ยกเว้น เมื่อ K = 5 และ ρ = 0.2 ดังตารางที่ 4.34

ตารางที่ 4.35 ค่าเฉลี่ย (ส่วนเบี่ยงเบนมาตรฐาน) จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ III

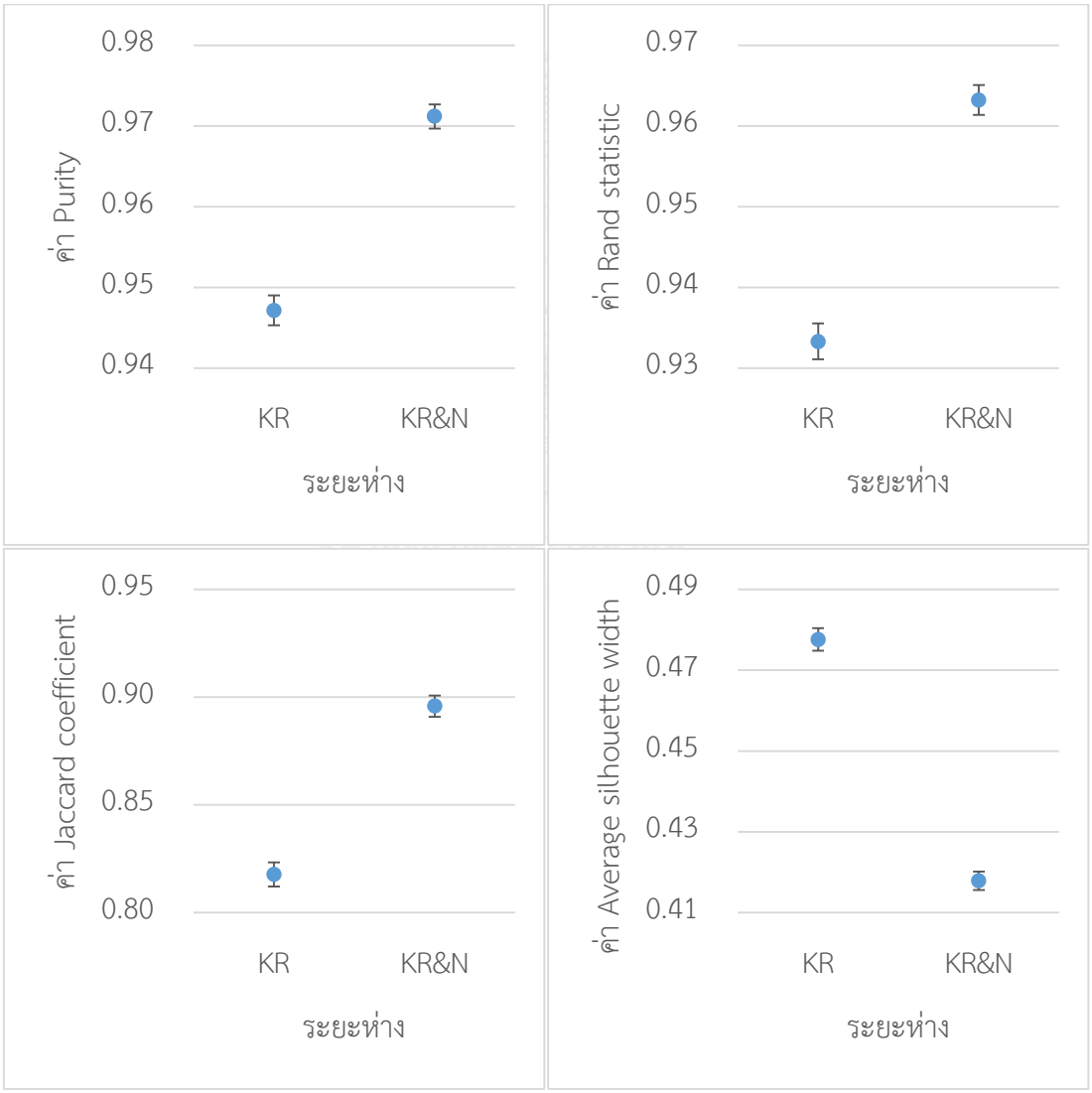
K	ρ	n	ค่าที่ได้	ระยะห่าง	
				KR	KR&N
3	0.2	20	ค่า Purity	0.9472 (0.0297)	**0.9712** (0.0242)
			ค่า Rand statistic	0.9333 (0.0359)	**0.9632** (0.0298)
			ค่า Jaccard coefficient	0.8177 (0.0899)	**0.8958** (0.0799)
			ค่า Average silhouette width	**0.4776** (0.0450)	0.4179 (0.0363)
		100	ค่า Purity	0.9473 (0.0131)	**0.9783** (0.0096)
			ค่า Rand statistic	0.9329 (0.0159)	**0.9718** (0.0121)
			ค่า Jaccard coefficient	0.8171 (0.0394	**0.9190** (0.0333)
			ค่า Average silhouette width	**0.4776** (0.0199)	0.3987 (0.0150)
	0.8	20	ค่า Purity	0.8512 (0.0471)	**0.8818** (0.0483)
			ค่า Rand statistic	0.8269 (0.0482)	**0.8621** (0.0495)
			ค่า Jaccard coefficient	0.5867 (0.0918)	**0.6576** (0.1018)
			ค่า Average silhouette width	**0.5303** (0.0580)	0.4532 (0.0465)

ตารางที่ 4.35 (ต่อ) ค่าเฉลี่ย (ส่วนเบี่ยงเบนมาตรฐาน) จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ III

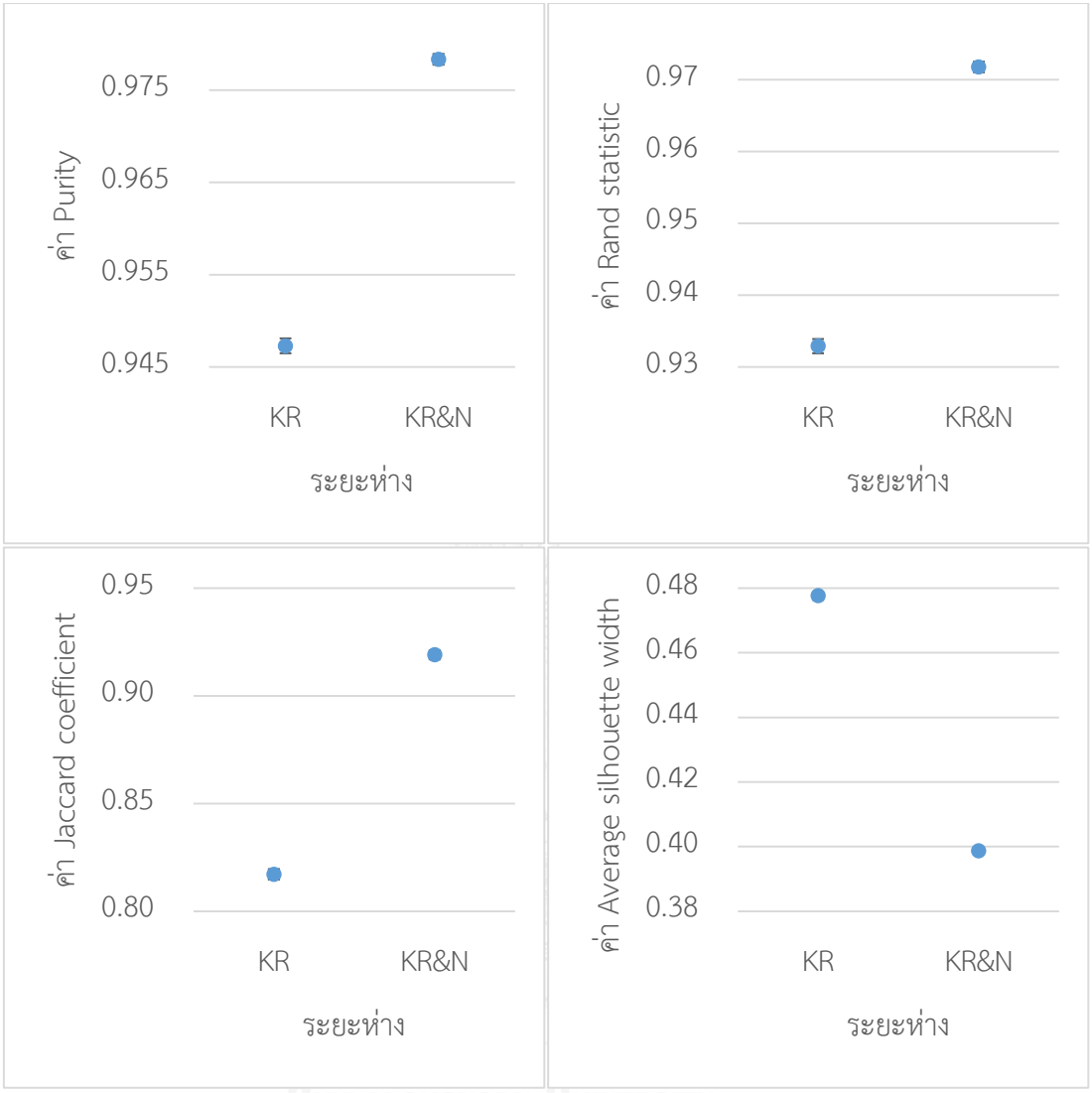
K	ρ	n	ค่าที่ได้	ระยะห่าง	
				KR	KR&N
3	0.8	100	ค่า Purity	0.8498 (0.0202)	**0.8854** (0.0239)
			ค่า Rand statistic	0.8231 (0.0207)	**0.8642** (0.0255)
			ค่า Jaccard coefficient	0.5821 (0.0383)	**0.6625** (0.0515)
			ค่า Average silhouette width	**0.5324** (0.0245)	0.4290 (0.0200)
5	0.2	20	ค่า Purity	**0.9160** (0.0284)	0.9141 (0.0290)
			ค่า Rand statistic	**0.9376** (0.0198)	0.9364 (0.0202)
			ค่า Jaccard coefficient	**0.7250** (0.0755)	0.7207 (0.0765)
			ค่า Average silhouette width	0.4586 (0.0363)	**0.4588** (0.0360)
		100	ค่า Purity	0.9123 (0.0134)	**0.9138** (0.0133)
			ค่า Rand statistic	0.9343 (0.0094)	**0.9354** (0.0093)
			ค่า Jaccard coefficient	0.7168 (0.0347)	**0.7208** (0.0347)
			ค่า Average silhouette width	0.4600 (0.0166)	**0.4608** (0.0167)
	0.8	20	ค่า Purity	0.7905 (0.0404)	**0.7920** (0.0397)
			ค่า Rand statistic	0.8572 (0.0235)	**0.8581** (0.0231)
			ค่า Jaccard coefficient	0.4648 (0.0643)	**0.4681** (0.0633)
			ค่า Average silhouette width	**0.5552** (0.0440)	0.5475 (0.0448)
		100	ค่า Purity	0.7883 (0.0182)	**0.7884** (0.0181)
			ค่า Rand statistic	0.8542 (0.0106)	**0.8543** (0.0105)
			ค่า Jaccard coefficient	0.4642 (0.0284)	**0.4644** (0.0282)
			ค่า Average silhouette width	**0.5597** (0.0198)	0.5567 (0.0200)

เมื่อพิจารณากราฟช่วงความเชื่อมั่น 95% ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ III พบว่า

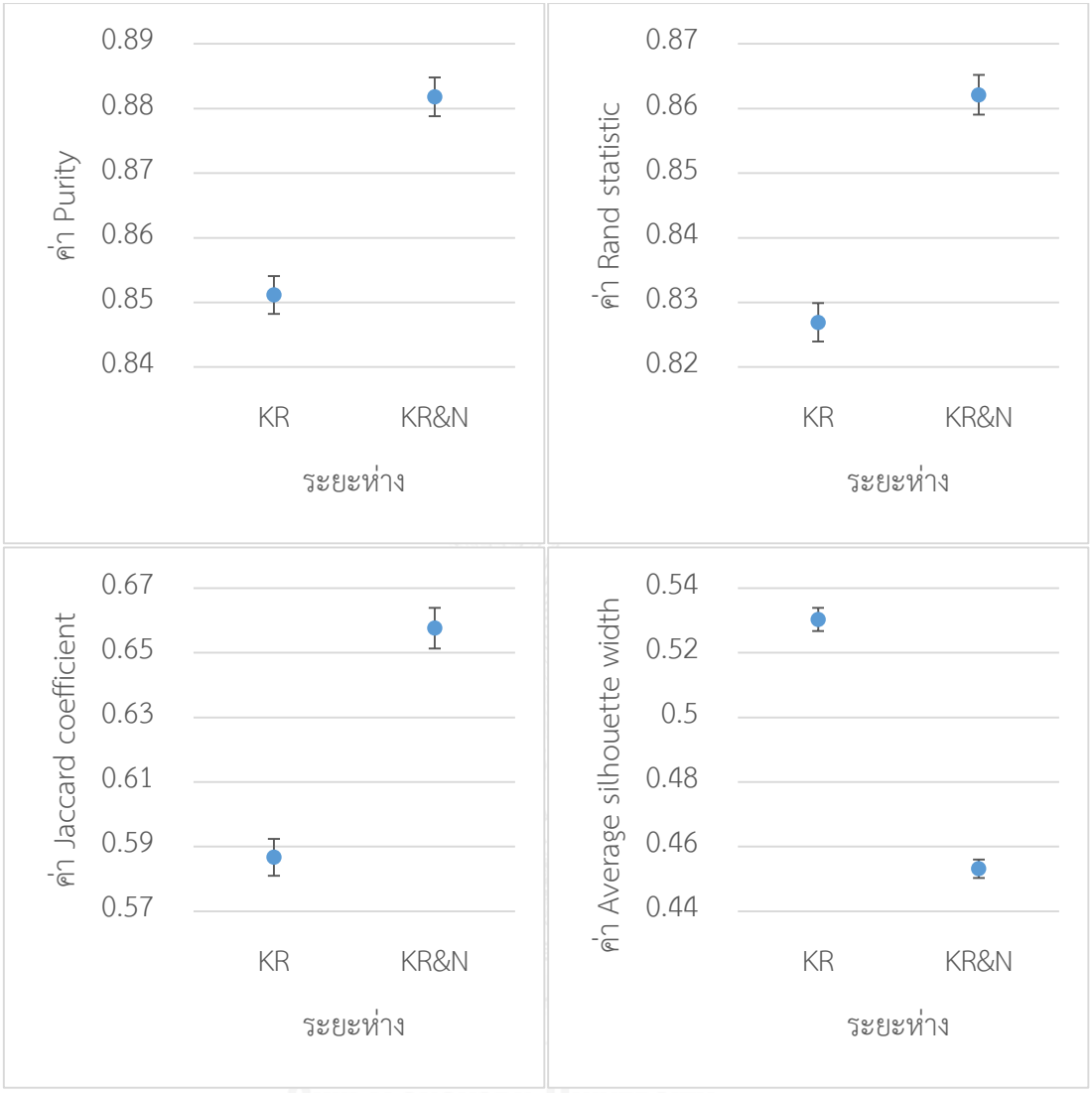
เมื่อ $K = 3$ พบว่า โดยเฉลี่ยค่า Purity ค่า Rand statistic และค่า Jaccard coefficient จากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR มีค่ามากกว่าจากการวิเคราะห์กลุ่มด้วยระยะห่างแบบ KR&N โดยไม่มีส่วนที่ทับซ้อนกันของช่วงความเชื่อมั่น 95% ในทุกกรณี อย่างไรก็ตามค่าเฉลี่ยของค่า Average silhouette width จากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR มีค่ามากกว่าจากการวิเคราะห์กลุ่มด้วยระยะห่างแบบ KR&N ดังรูปที่ 4.17 ถึง 4.20



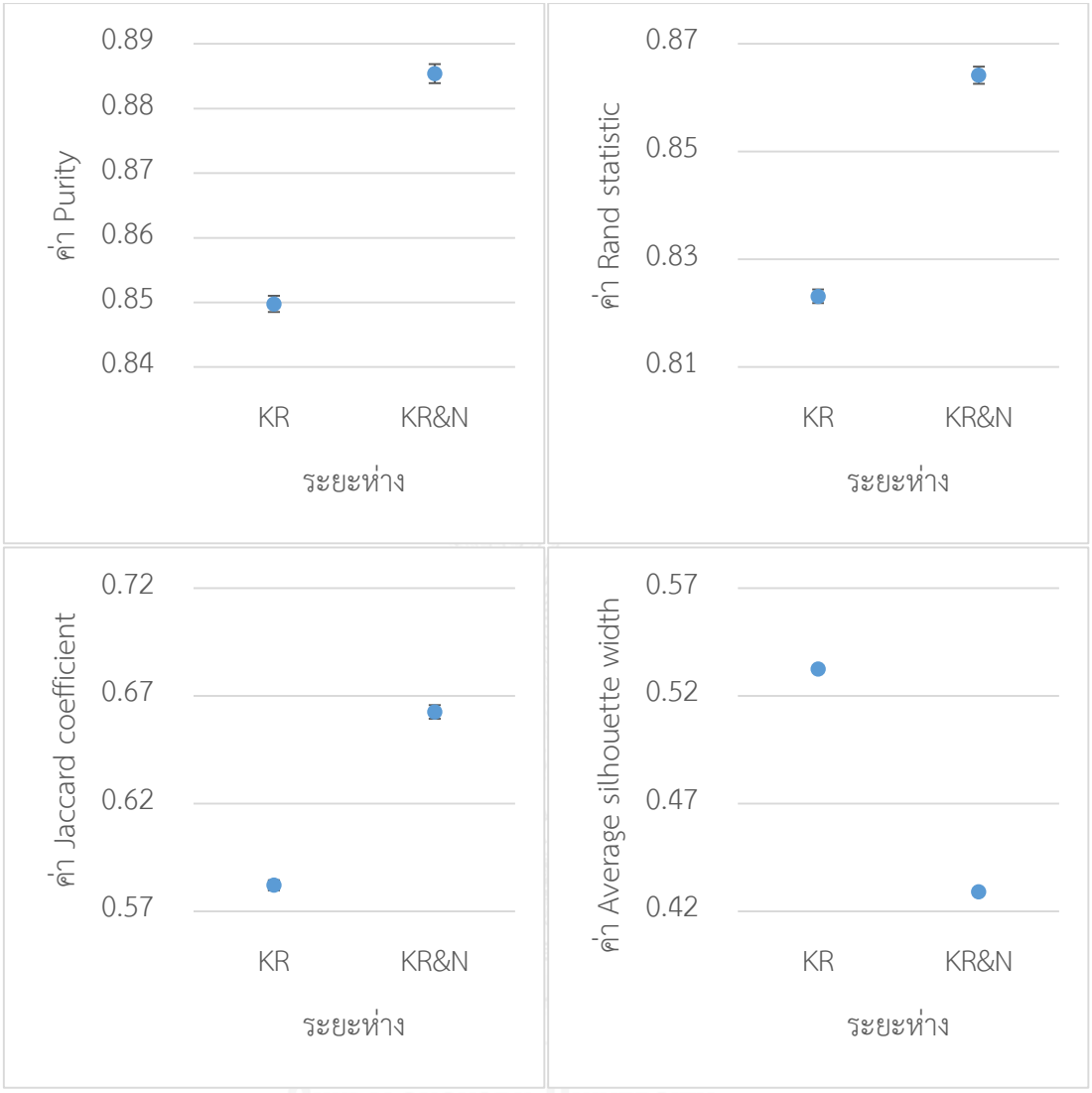
รูปที่ 4.17 กราฟช่วงความเชื่อมั่น 95% ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ III เมื่อ $K = 3$, $\rho = 0.2$ และ $n = 20$



รูปที่ 4.18 กราฟช่วงความเชื่อมั่น 95% ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ III เมื่อ $K = 3$, $\rho = 0.2$ และ $n = 100$



รูปที่ 4.19 กราฟช่วงความเชื่อมั่น 95% ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ III เมื่อ $K = 3$, $\rho = 0.8$ และ $n = 20$

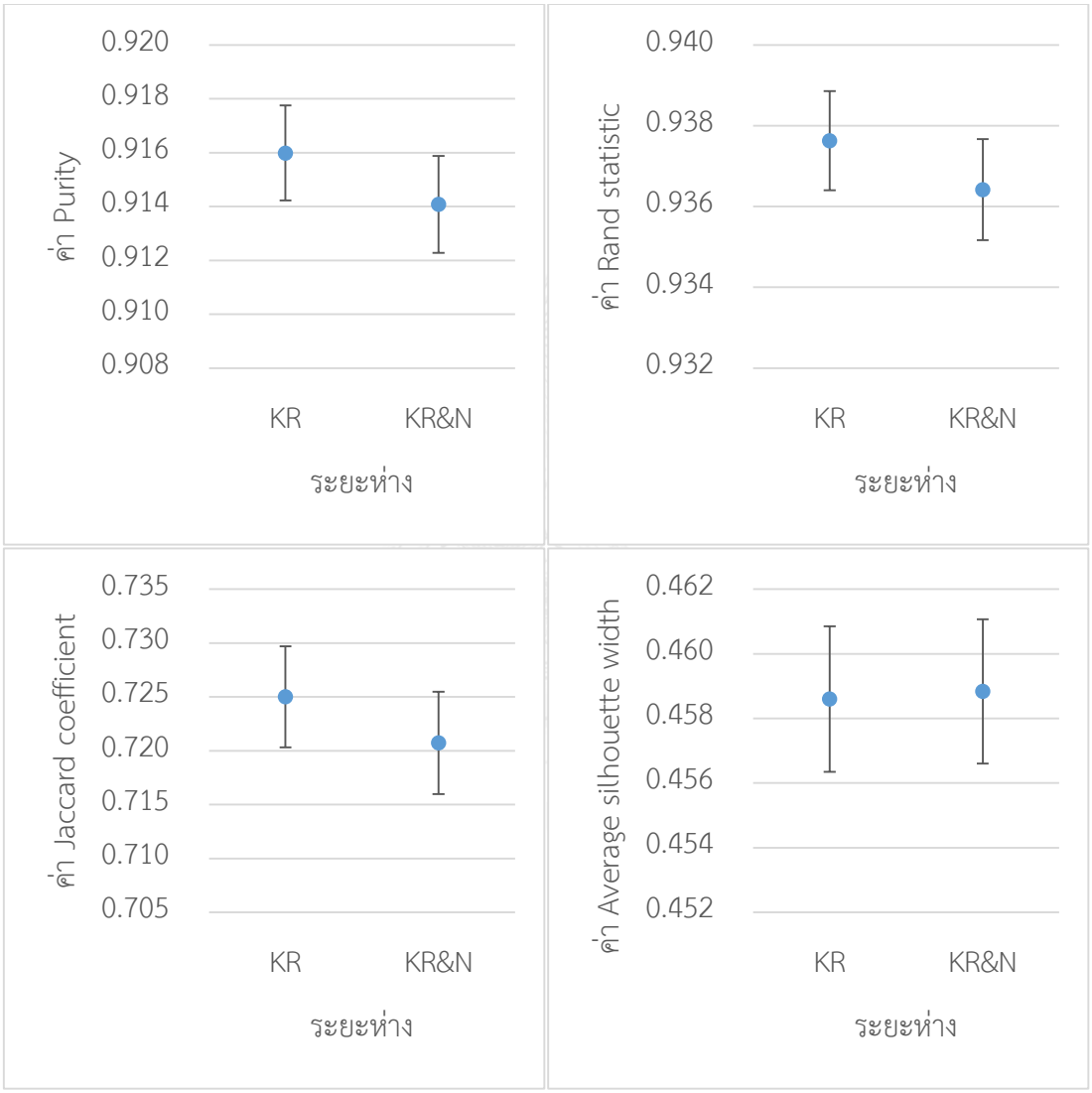


รูปที่ 4.20 กราฟช่วงความเชื่อมั่น 95% ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ III เมื่อ $K = 3$, $\rho = 0.8$ และ $n = 100$

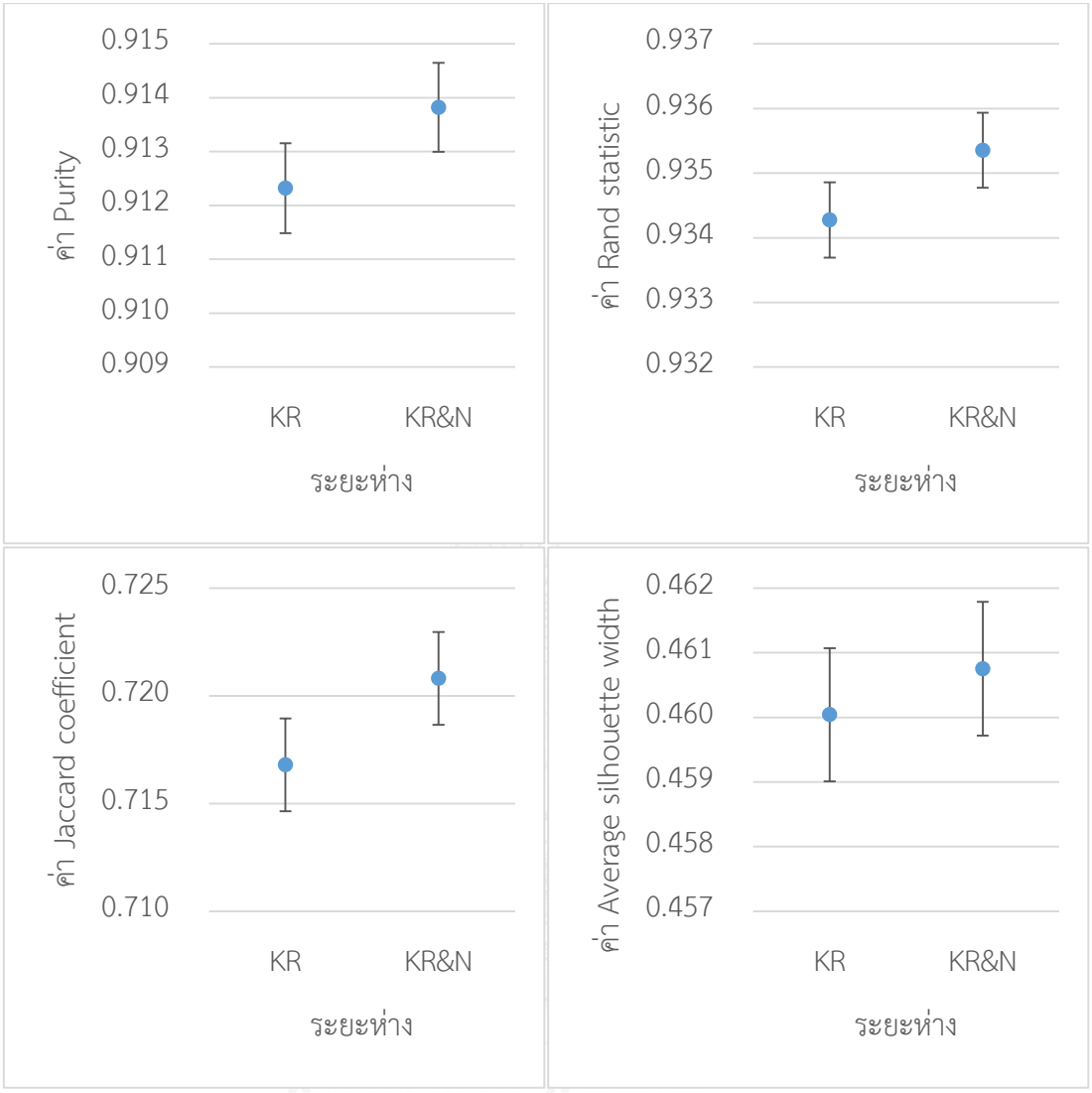
เมื่อ $K = 5$ $\rho = 0.2$ และ $n = 20$ การวิเคราะห์กลุ่มด้วยระยะห่างของ KR มีประสิทธิภาพดีกว่าระยะห่างแบบ KR&N โดยเฉลี่ยอย่างใกล้เคียงกัน โดยที่มีช่วงความเชื่อมั่น 95% ทับซ้อนกันอยู่ ดังรูปที่ 4.21 ขณะที่ในกรณีอื่น ๆ ที่ $K = 5$ ให้ผลตรงกันข้ามกัน คือ การวิเคราะห์กลุ่มด้วยระยะห่างแบบ KR&N มีประสิทธิภาพดีกว่าระยะห่างของ KR โดยเฉลี่ย แต่ยังคงมีส่วนของช่วงความเชื่อมั่น 95% ทับซ้อนกันอยู่ ทั้งนี้เมื่อค่าสัมประสิทธิ์สหสัมพันธ์เพิ่มขึ้นจาก $\rho = 0.2$ เป็น $\rho = 0.8$ พบว่าช่วงความเชื่อมั่น 95% ซ้อนทับกันมากยิ่งขึ้น ดังรูปที่ 4.22 ถึง 4.24

เมื่อพิจารณาค่าเฉลี่ยของค่า Average silhouette width พบว่าเมื่อ $K = 5$ และ $\rho = 0.2$ ค่าเฉลี่ยของค่า Average silhouette width จากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR และ

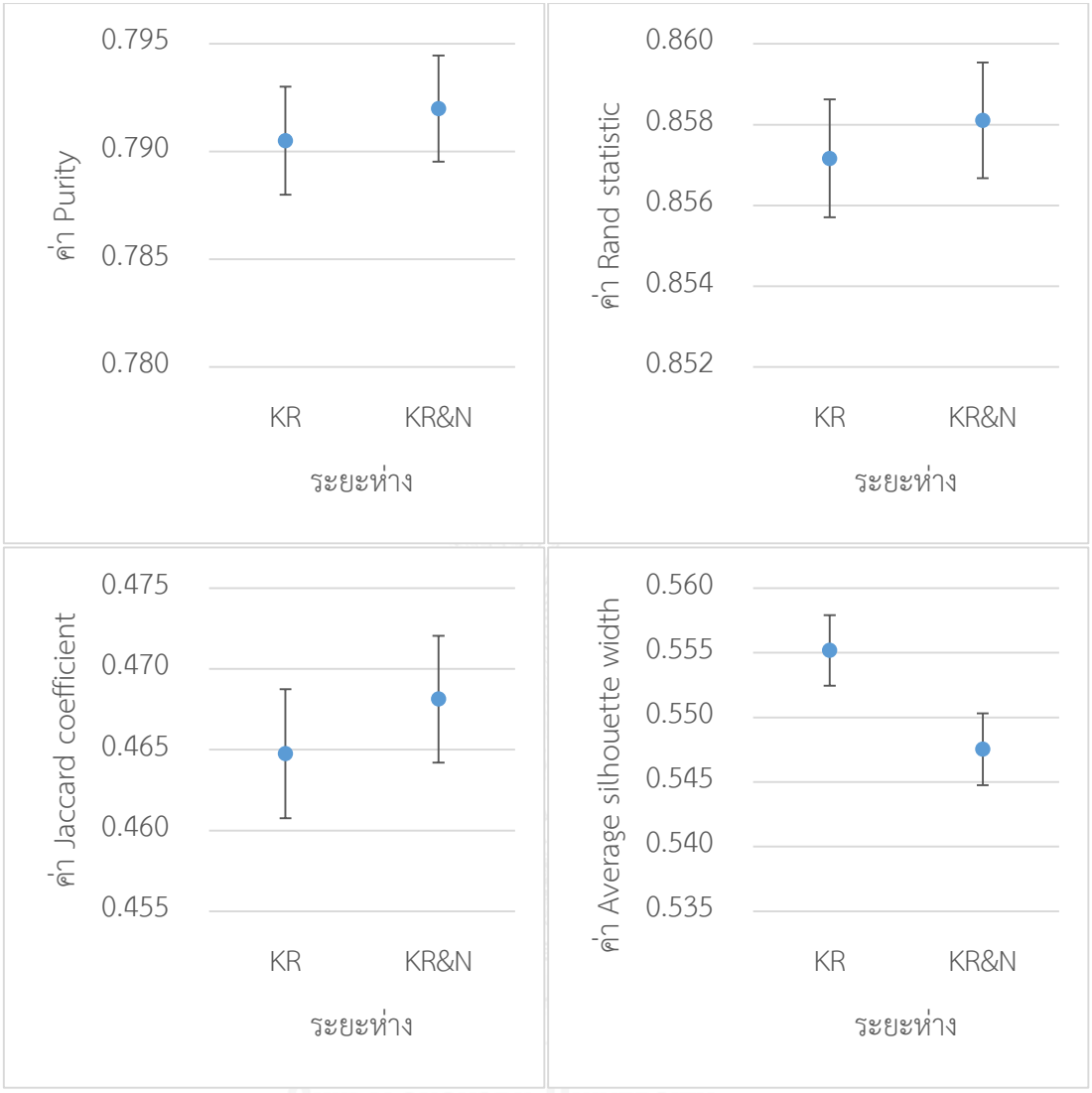
ระยะห่างแบบ KR&N ใกล้เคียงกัน ดังรูปที่ 4.21 และ 4.22 แต่เมื่อ $K = 5$ และ $\rho = 0.8$ พบว่าค่าเฉลี่ยของค่า Average silhouette width จากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR มีค่ามากกว่าจากการวิเคราะห์กลุ่มด้วยระยะห่างแบบ KR&N โดยไม่มีส่วนที่ทับซ้อนกันของช่วงความเชื่อมั่น 95% ดังรูปที่ 4.23 และ 4.24



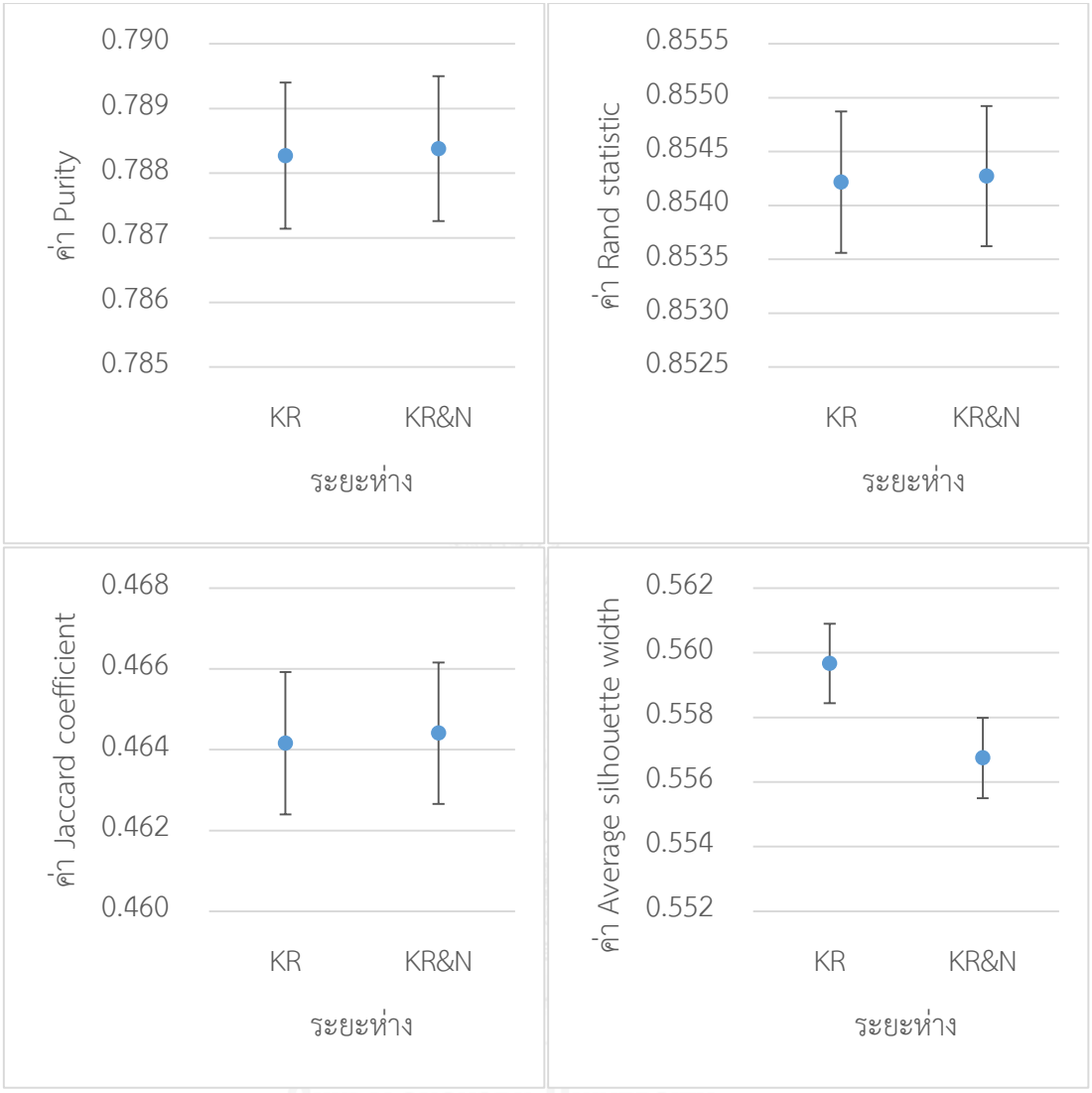
รูปที่ 4.21 กราฟช่วงความเชื่อมั่น 95% ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ III เมื่อ $K = 5$, $\rho = 0.2$ และ $n = 20$



รูปที่ 4.22 กราฟช่วงความเชื่อมั่น 95% ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ III เมื่อ $K = 5$, $\rho = 0.2$ และ $n = 100$



รูปที่ 4.23 กราฟช่วงความเชื่อมั่น 95% ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ III เมื่อ $K = 5$, $\rho = 0.8$ และ $n = 20$



รูปที่ 4.24 กราฟช่วงความเชื่อมั่น 95% ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ III เมื่อ $K = 5$, $\rho = 0.8$ และ $n = 100$

เนื่องจากการพิจารณากราฟช่วงความเชื่อมั่น 95% พบว่าค่าเฉลี่ยของค่า Purity ค่า Rand statistic ค่า Jaccard coefficient และค่า Average silhouette width มีค่าใกล้เคียงกัน ทั้งยังมีส่วนของช่วงความเชื่อมั่น 95% ทับซ้อนกัน จึงไม่สามารถสรุปได้ว่าการวิเคราะห์กลุ่มด้วยระยะห่างแบบใดมีประสิทธิภาพดีที่สุด ผู้วิจัยจึงทดสอบความแตกต่างระหว่างค่าเฉลี่ยของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแต่ละคู่ ด้วยสถิติทดสอบ t และแบ่งการพิจารณาข้อมูลรูปแบบที่ III เป็นกรณีใหญ่ 2 กรณีตามจำนวนกลุ่มข้อมูล คือกรณีที่ $K = 3$ และ $K = 5$

4.1.3.1 ผลทดสอบความแตกต่างการวิเคราะห์กลุ่มข้อมูลรูปแบบที่ III เมื่อ K = 3

เมื่อ K = 3 ข้อมูลยังถูกแบ่งออกเป็นกรณีต่าง ๆ ตามค่าสัมประสิทธิ์สหสัมพันธ์และขนาดข้อมูลต่อกลุ่ม จึงได้ข้อมูลที่แตกต่างกันดังต่อไปนี้

ข้อมูลรูปแบบที่ III เมื่อ $K = 3,\ \rho = 0.2$ *และ* $n = 20$

ข้อมูลรูปแบบที่ III เมื่อ $K = 3,\ \rho = 0.2$ *และ* $n = 100$

ข้อมูลรูปแบบที่ III เมื่อ $K = 3,\ \rho = 0.8$ *และ* $n = 20$

ข้อมูลรูปแบบที่ III เมื่อ $K = 3,\ \rho = 0.8$ *และ* $n = 100$

ทำการศึกษาประสิทธิภาพการวิเคราะห์กลุ่มข้อมูลต่าง ๆ โดยทดสอบความแตกต่างระหว่างค่าเฉลี่ยของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างระหว่างระยะห่างของ KR และระยะห่างแบบ KR&N ด้วยค่าสถิติทดสอบ t พบว่าในทุกกรณี ค่า Purity ค่า Rand statistic ค่า Jaccard coefficient และค่า Average silhouette width จากการวิเคราะห์กลุ่ม มี Sig. น้อยกว่า 0.05 จึงปฏิเสธสมมติฐานหลัก ดังนั้น การวิเคราะห์กลุ่มด้วยระยะห่างของ KR มีประสิทธิภาพและค่า Average silhouette width แตกต่างจากการวิเคราะห์กลุ่มด้วยระยะห่างแบบ KR&N ที่ระดับนัยสำคัญ 0.05 ดังตารางที่ 4.36 ถึง 4.39

ตารางที่ 4.36 ผลทดสอบความแตกต่างระหว่างค่าเฉลี่ยของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR และระยะห่างแบบ KR&N สำหรับข้อมูลรูปแบบที่ III เมื่อ K = 3, ρ = 0.2 และ n = 20 ด้วยสถิติทดสอบ t

ค่าที่วัด	t	df	Sig. (2-tailed)	ผลต่างค่าเฉลี่ย	ผลต่าง Std. Error
ค่า Purity	-19.8074	1918.2074	**0.0000**	-0.0240	0.0012
ค่า Rand statistic	-20.2633	1931.5452	**0.0000**	-0.0299	0.0015
ค่า Jaccard coefficient	-20.5424	1970.8957	**0.0000**	-0.0781	0.0038
ค่า Average silhouette width	32.6392	1912.8781	**0.0000**	0.0597	0.0018

ตารางที่ 4.37 ผลทดสอบความแตกต่างระหว่างค่าเฉลี่ยของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR และระยะห่างแบบ KR&N สำหรับข้อมูลรูปแบบที่ III เมื่อ K = 3, ρ = 0.2 และ n = 100 ด้วยสถิติทดสอบ t

ค่าที่วัด	t	df	Sig. (2-tailed)	ผลต่างค่าเฉลี่ย	ผลต่าง Std. Error
ค่า Purity	-60.5793	1830.4981	**0.0000**	-0.0310	0.0005
ค่า Rand statistic	-61.4878	1867.5758	**0.0000**	-0.0389	0.0006
ค่า Jaccard coefficient	-62.4870	1944.2278	**0.0000**	-0.1018	0.0016
ค่า Average silhouette width	100.0984	1859.2957	**0.0000**	0.0789	0.0008

ตารางที่ 4.38 ผลทดสอบความแตกต่างระหว่างค่าเฉลี่ยของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR และระยะห่างแบบ KR&N สำหรับข้อมูลรูปแบบที่ III เมื่อ K = 3, ρ = 0.8 และ n = 20 ด้วยสถิติทดสอบ t

ค่าที่วัด	t	df	Sig. (2-tailed)	ผลต่างค่าเฉลี่ย	ผลต่าง Std. Error
ค่า Purity	-14.3572	1998.0000	**0.0000**	-0.0306	0.0021
ค่า Rand statistic	-16.1142	1998.0000	**0.0000**	-0.0352	0.0022
ค่า Jaccard coefficient	-16.3575	1977.0935	**0.0000**	-0.0709	0.0043
ค่า Average silhouette width	32.8046	1908.3475	**0.0000**	0.0771	0.0024

ตารางที่ 4.39 ผลทดสอบความแตกต่างระหว่างค่าเฉลี่ยของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR และระยะห่างแบบ KR&N สำหรับข้อมูลรูปแบบที่ III เมื่อ K = 3, ρ = 0.8 และ n = 100 ด้วยสถิติทดสอบ t

ค่าที่วัด	t	df	Sig. (2-tailed)	ผลต่างค่าเฉลี่ย	ผลต่าง Std. Error
ค่า Purity	-35.9928	1943.5917	**0.0000**	-0.0356	0.0010
ค่า Rand statistic	-39.4257	1917.7126	**0.0000**	-0.0410	0.0010
ค่า Jaccard coefficient	-39.6400	1845.5179	**0.0000**	-0.0805	0.0020
ค่า Average silhouette width	103.4016	1920.9580	**0.0000**	0.1034	0.0010

4.1.3.2 ผลทดสอบความแตกต่างการวิเคราะห์กลุ่มข้อมูลรูปแบบที่ III เมื่อ K = 5

เมื่อ K = 5 ข้อมูลยังถูกแบ่งออกเป็นกรณีต่าง ๆ ตามค่าสัมประสิทธิ์สหสัมพันธ์และขนาดข้อมูลต่อกลุ่ม และมีผลการทดสอบความแตกต่างของประสิทธิภาพการวิเคราะห์กลุ่มข้อมูลด้วยสถิติทดสอบ t ดังต่อไปนี้

ข้อมูลรูปแบบที่ III เมื่อ K = 5, ρ = 0.2 และ n = 20

การวิเคราะห์กลุ่มด้วยระยะห่างของ KR มีประสิทธิภาพดีกว่าระยะห่างแบบ KR&N โดยเฉลี่ย แต่เมื่อทดสอบความแตกต่างของค่าเฉลี่ยของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างทั้ง 2 วิธีนี้ ด้วยค่าสถิติทดสอบ t พบว่า ค่า Purity ค่า Rand statistic ค่า Jaccard coefficient และค่า Average silhouette width จากการวิเคราะห์กลุ่ม มี Sig. (2-tailed) มากกว่ากว่า 0.05 ดังตารางที่ 4.40 จึงไม่สามารถปฏิเสธสมมติฐานหลักได้ ดังนั้นการวิเคราะห์กลุ่มด้วยระยะห่างของ KR มีประสิทธิภาพและมีค่า Average silhouette width ไม่แตกต่างจากการวิเคราะห์กลุ่มด้วยระยะห่างแบบ KR&N

ตารางที่ 4.40 ผลทดสอบความแตกต่างระหว่างค่าเฉลี่ยของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR และระยะห่างแบบ KR&N สำหรับข้อมูลรูปแบบที่ III เมื่อ K = 5, ρ = 0.2 และ n = 20 ด้วยสถิติทดสอบ t

ค่าที่วัด	t	df	Sig. (2-tailed)	ผลต่างค่าเฉลี่ย	ผลต่าง Std. Error
ค่า Purity	1.4886	1998.0000	0.1367	0.0019	0.0013
ค่า Rand statistic	1.3578	1998.0000	0.1747	0.0012	0.0009
ค่า Jaccard coefficient	1.2588	1998.0000	0.2082	0.0043	0.0034
ค่า Average silhouette width	-0.1455	1998.0000	0.8843	-0.0002	0.0016

ข้อมูลรูปแบบที่ III เมื่อ K = 5, ρ = 0.2 และ n = 100

กรณีที่ K = 5, ρ = 0.2 เมื่อจำนวนข้อมูลต่อกลุ่มเพิ่มขึ้นจาก n = 20 เป็น n = 100 ดังตารางที่ 4.35 การวิเคราะห์กลุ่มด้วยระยะห่างแบบ KR&N กลับมีประสิทธิภาพดีกว่าระยะห่างของ KR จึงทดสอบความแตกต่างระหว่างค่าเฉลี่ยของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างระหว่างระยะห่างของ KR และระยะห่างแบบ KR&N ด้วยค่าสถิติทดสอบ t ดังตารางที่ 4.41 ทำให้ทราบว่าค่า Purity ค่า Rand statistic ค่า Jaccard coefficient จากการวิเคราะห์กลุ่ม มี Sig. (2-tailed) น้อยกว่า 0.05 จึงปฏิเสธสมมติฐานหลัก ดังนั้นการวิเคราะห์กลุ่มด้วยระยะห่างของ KR มีประสิทธิภาพแตกต่างจากการวิเคราะห์กลุ่มด้วยระยะห่างแบบ KR&N ที่ระดับนัยสำคัญ 0.05 ขณะที่ค่า Average silhouette width มี Sig. (2-tailed) มากกว่า 0.05 จึงไม่สามารถปฏิเสธสมมติฐานหลักได้ ดังนั้นการวิเคราะห์กลุ่มด้วยระยะห่างของ KR มีค่า Average silhouette width ไม่แตกต่างจากการวิเคราะห์กลุ่มด้วยระยะห่างแบบ KR&N

ตารางที่ 4.41 ผลทดสอบความแตกต่างระหว่างค่าเฉลี่ยของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR และระยะห่างแบบ KR&N สำหรับข้อมูลรูปแบบที่ III เมื่อ K = 5, ρ = 0.2 และ n = 100 ด้วยสถิติทดสอบ t

ค่าที่วัด	t	df	Sig. (2-tailed)	ผลต่างค่าเฉลี่ย	ผลต่าง Std. Error
ค่า Purity	-2.5072	1998.0000	**0.0122**	-0.0015	0.0006
ค่า Rand statistic	-2.5728	1998.0000	**0.0102**	-0.0011	0.0004
ค่า Jaccard coefficient	-2.5860	1998.0000	**0.0098**	-0.0040	0.0016
ค่า Average silhouette width	-0.9573	1998.0000	0.3385	-0.0007	0.0007

การทดสอบความแตกต่างระหว่างค่าเฉลี่ยของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR และระยะห่างแบบ KR&N ด้วยสถิติทดสอบ t สำหรับข้อมูลในกรณีต่อไปนี้

ข้อมูลรูปแบบที่ III เมื่อ $K = 5,\ \rho = 0.8$ *และ* $n = 20$

ข้อมูลรูปแบบที่ III เมื่อ $K = 5,\ \rho = 0.8$ *และ* $n = 100$

พบว่า ค่า Purity ค่า Rand statistic และค่า Jaccard coefficient จากการวิเคราะห์กลุ่ม มี Sig. (2-tailed) มากกว่า 0.05 จึงไม่สามารถปฏิเสธสมมติฐานหลัก ดังนั้นการวิเคราะห์กลุ่มด้วยระยะห่างของ KR มีประสิทธิภาพไม่แตกต่างจากการวิเคราะห์กลุ่มด้วยระยะห่างแบบ KR&N ขณะที่ค่า Average silhouette width มี Sig. (2-tailed) น้อยกว่า 0.05 จึงปฏิเสธสมมติฐานหลัก ดังนั้นการวิเคราะห์กลุ่มด้วยระยะห่างของ KR มีค่า Average silhouette width แตกต่างจากการวิเคราะห์กลุ่มด้วยระยะห่างแบบ KR&N ที่ระดับนัยสำคัญ 0.05 ดังตารางที่ 4.42 และ 4.43

ตารางที่ 4.42 ผลทดสอบความแตกต่างระหว่างค่าเฉลี่ยของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR และระยะห่างแบบ KR&N สำหรับข้อมูลรูปแบบที่ III เมื่อ $K = 5,\ \rho = 0.8$ และ $n = 20$ ด้วยสถิติทดสอบ t

ค่าที่วัด	t	df	Sig. (2-tailed)	ผลต่างค่าเฉลี่ย	ผลต่าง Std. Error
ค่า Purity	-0.8324	1998.0000	0.4053	-0.0015	0.0018
ค่า Rand statistic	-0.9027	1998.0000	0.3668	-0.0009	0.0010
ค่า Jaccard coefficient	-1.1811	1998.0000	0.2377	-0.0034	0.0029
ค่า Average silhouette width	3.8487	1998.0000	**0.0001**	0.0076	0.0020

ตารางที่ 4.43 ผลทดสอบความแตกต่างระหว่างค่าเฉลี่ยของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR และระยะห่างแบบ KR&N สำหรับข้อมูลรูปแบบที่ III เมื่อ $K = 5,\ \rho = 0.8$ และ $n = 100$ ด้วยสถิติทดสอบ t

ค่าที่วัด	t	df	Sig. (2-tailed)	ผลต่างค่าเฉลี่ย	ผลต่าง Std. Error
ค่า Purity	-0.1306	1997.8624	0.8961	-0.0001	0.0008
ค่า Rand statistic	-0.1214	1997.8724	0.9034	-0.0001	0.0005
ค่า Jaccard coefficient	-0.1973	1997.8719	0.8436	-0.0002	0.0013
ค่า Average silhouette width	3.2791	1998.0000	**0.0011**	0.0029	0.0009

4.1.4 ผลการวิเคราะห์กลุ่มข้อมูลรูปแบบที่ IV

ผู้วิจัยทำการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ 2 วิธี ได้แก่ ระยะห่างของ KR และระยะห่างของ P สำหรับข้อมูลรูปแบบที่ IV เนื่องจากเป็นข้อมูลที่ประกอบไปด้วยตัวแปรอันดับและตัวแปรเชิงปริมาณ อย่างละ 3 ตัวแปร และคำนวณค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของ ค่า Purity ค่า Rand statistic ค่า Jaccard coefficient และค่า Average silhouette width ได้ผลดังตารางที่ 4.44

ผลการเปรียบเทียบประสิทธิภาพโดยเฉลี่ยในการวิเคราะห์กลุ่มด้วยมาตรวัดระยะห่างแบบต่าง ๆ พบว่า เมื่อ K = 3 การวิเคราะห์กลุ่มด้วยระยะห่างของ KR ให้ค่าเฉลี่ยของค่า Purity ค่า Rand statistic และค่า Jaccard coefficient สูงที่สุดทุกกรณี แสดงว่าการวิเคราะห์กลุ่มด้วยระยะห่างของ KR มีประสิทธิภาพดีกว่าการวิเคราะห์กลุ่มด้วยระยะห่างของ P ซึ่งแตกต่างจากกรณีที่ K = 5 การวิเคราะห์กลุ่มด้วยระยะห่างของ P ให้ค่าเฉลี่ยของค่า Purity ค่า Rand statistic และค่า Jaccard coefficient สูงที่สุดทุกกรณี แสดงว่าการวิเคราะห์กลุ่มด้วยระยะห่างของ P มีประสิทธิภาพดีกว่าการวิเคราะห์กลุ่มด้วยระยะห่างของ KR ดังตารางที่ 4.44

นอกจากนี้ค่าเฉลี่ยของค่า Average silhouette width จากการวิเคราะห์กลุ่มด้วยระยะห่างของ P มีค่ามากกว่าจากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR ในทุกกรณี ดังตารางที่ 4.44

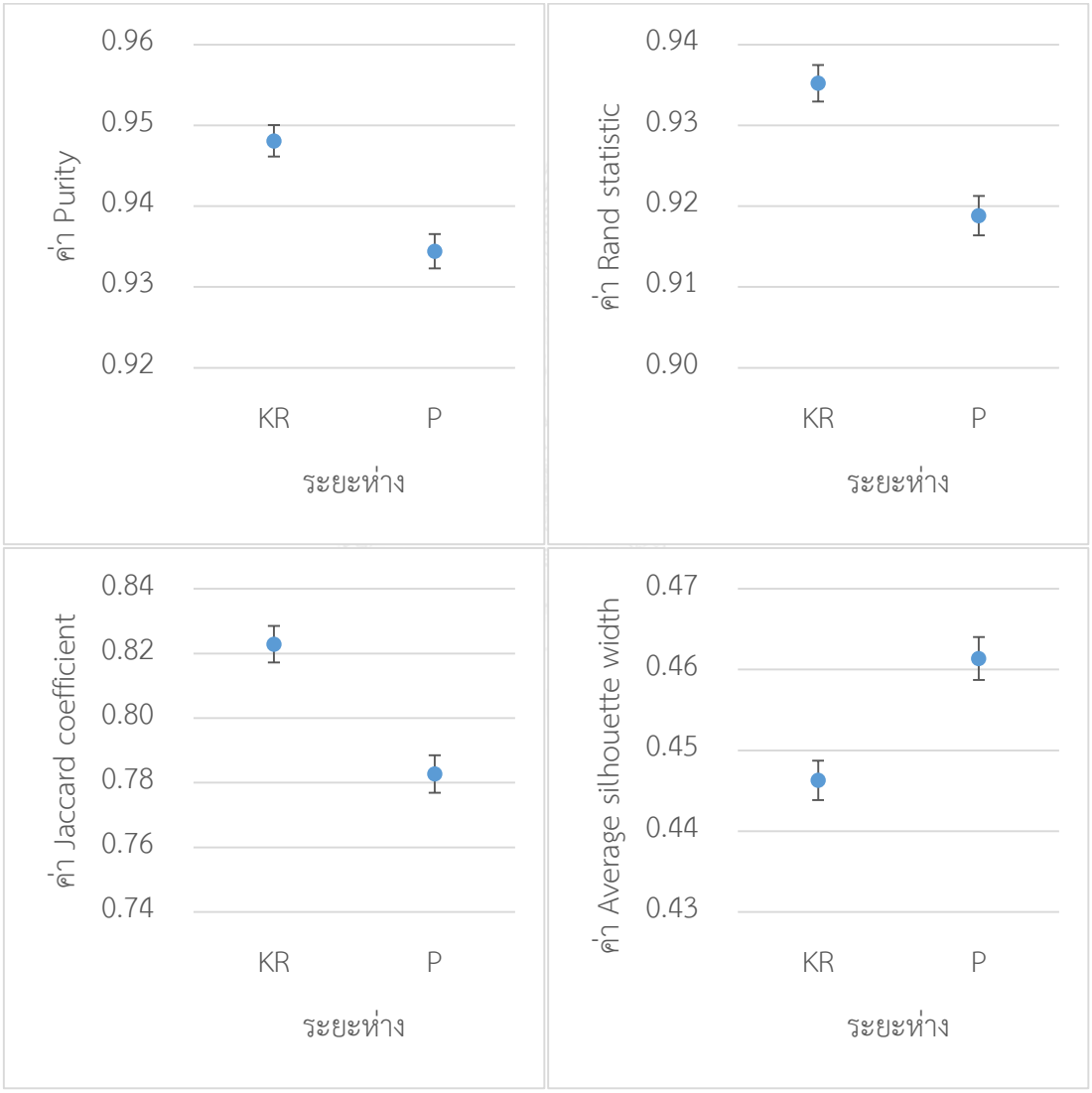
ตารางที่ 4.44 ค่าเฉลี่ย (ส่วนเบี่ยงเบนมาตรฐาน) จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ IV

K	ρ	n	ค่าที่วัด	ระยะห่าง	
				KR	P
3	0.2	20	ค่า Purity	**0.9481** (0.0313)	0.9344 (0.0343)
			ค่า Rand statistic	**0.9352** (0.0365)	0.9188 (0.0392)
			ค่า Jaccard coefficient	**0.8229** (0.0905)	0.7827 (0.0934)
			ค่า Average silhouette width	0.4463 (0.0392)	**0.4614** (0.0428)
		100	ค่า Purity	**0.9456** (0.0149)	0.9320 (0.0162)
			ค่า Rand statistic	**0.9316** (0.0177)	0.9148 (0.0189)
			ค่า Jaccard coefficient	**0.8142** (0.0435)	0.7739 (0.0442)
			ค่า Average silhouette width	0.4382 (0.0178)	**0.4545** (0.0194)

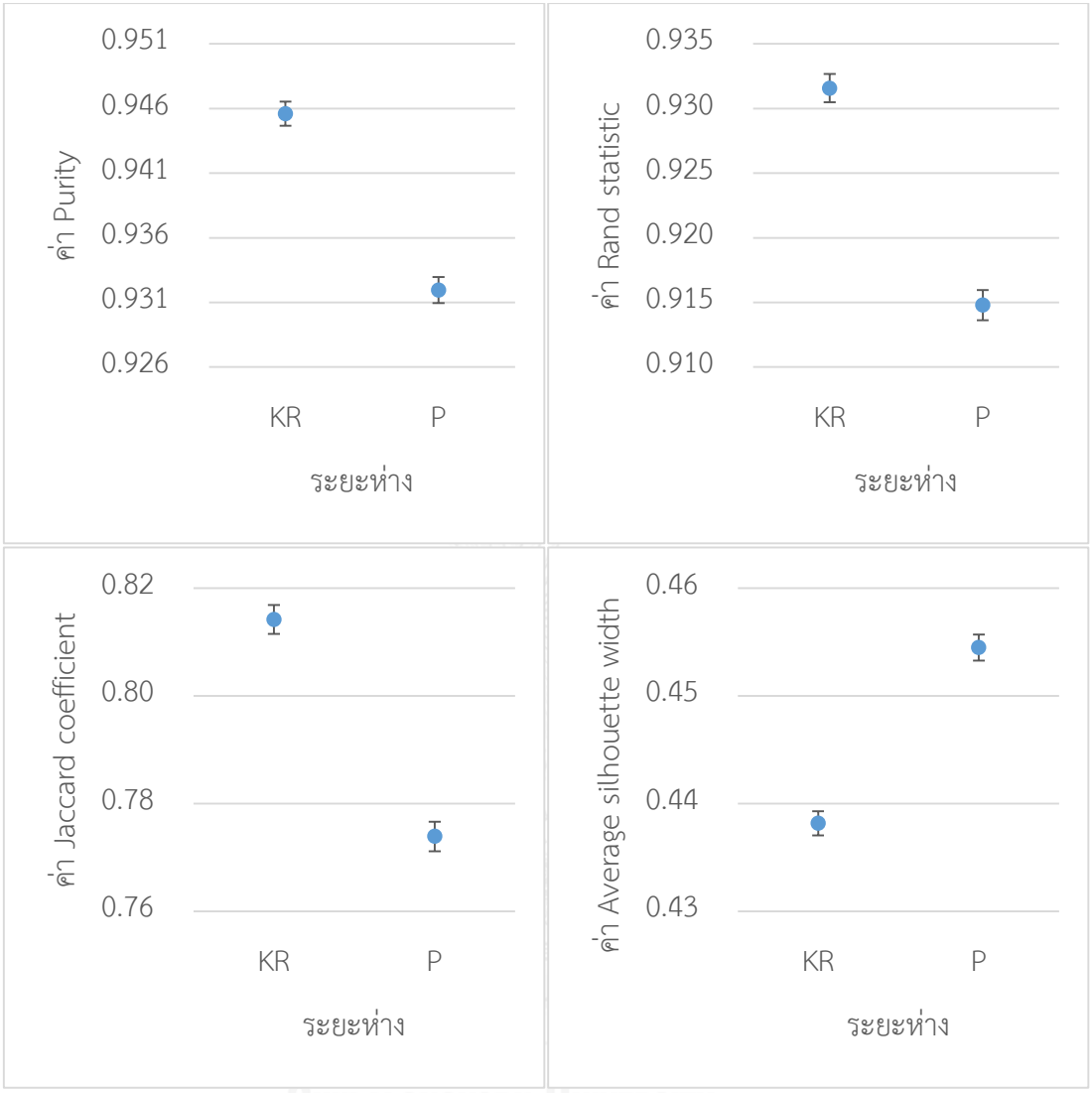
ตารางที่ 4.44 (ต่อ) ค่าเฉลี่ย (ส่วนเบี่ยงเบนมาตรฐาน) จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ IV

K	ρ	n	ค่าที่วัด	ระยะห่าง	
				KR	P
3	0.8	20	ค่า Purity	**0.8349** (0.0481)	0.8249 (0.0477)
			ค่า Rand statistic	**0.8162** (0.0438)	0.8009 (0.0456)
			ค่า Jaccard coefficient	**0.5665** (0.0801)	0.5402 (0.0795)
			ค่า Average silhouette width	0.5146 (0.0476)	**0.5424** (0.0519)
		100	ค่า Purity	**0.8303** (0.0230)	0.8168 (0.0224)
			ค่า Rand statistic	**0.8107** (0.0207)	0.7893 (0.0218)
			ค่า Jaccard coefficient	**0.5601** (0.0368)	0.5245 (0.0362)
			ค่า Average silhouette width	0.5127 (0.0221)	**0.5489** (0.0235)
5	0.2	20	ค่า Purity	0.9053 (0.0320)	**0.9063** (0.0323)
			ค่า Rand statistic	0.9305 (0.0219)	**0.9311** (0.0220)
			ค่า Jaccard coefficient	0.6997 (0.0792)	**0.7021** (0.0796)
			ค่า Average silhouette width	0.3651 (0.0348)	**0.3668** (0.0346)
		100	ค่า Purity	0.9069 (0.0142)	**0.9072** (0.0142)
			ค่า Rand statistic	0.9305 (0.0099)	**0.9306** (0.0099)
			ค่า Jaccard coefficient	0.7036 (0.0357)	**0.7041** (0.0357)
			ค่า Average silhouette width	0.3636 (0.0157)	**0.3639** (0.0157)
	0.8	20	ค่า Purity	0.7759 (0.0390)	**0.7773** (0.0392)
			ค่า Rand statistic	0.8489 (0.0223)	**0.8493** (0.0228)
			ค่า Jaccard coefficient	0.4442 (0.0585)	**0.4456** (0.0595)
			ค่า Average silhouette width	0.4525 (0.0438)	**0.4535** (0.0442)
		100	ค่า Purity	0.7685 (0.0188)	**0.7689** (0.0186)
			ค่า Rand statistic	0.8435 (0.0104)	**0.8436** (0.0103)
			ค่า Jaccard coefficient	0.4364 (0.0267)	**0.4368** (0.0265)
			ค่า Average silhouette width	**0.4548** (0.0189)	**0.4548** (0.0190)

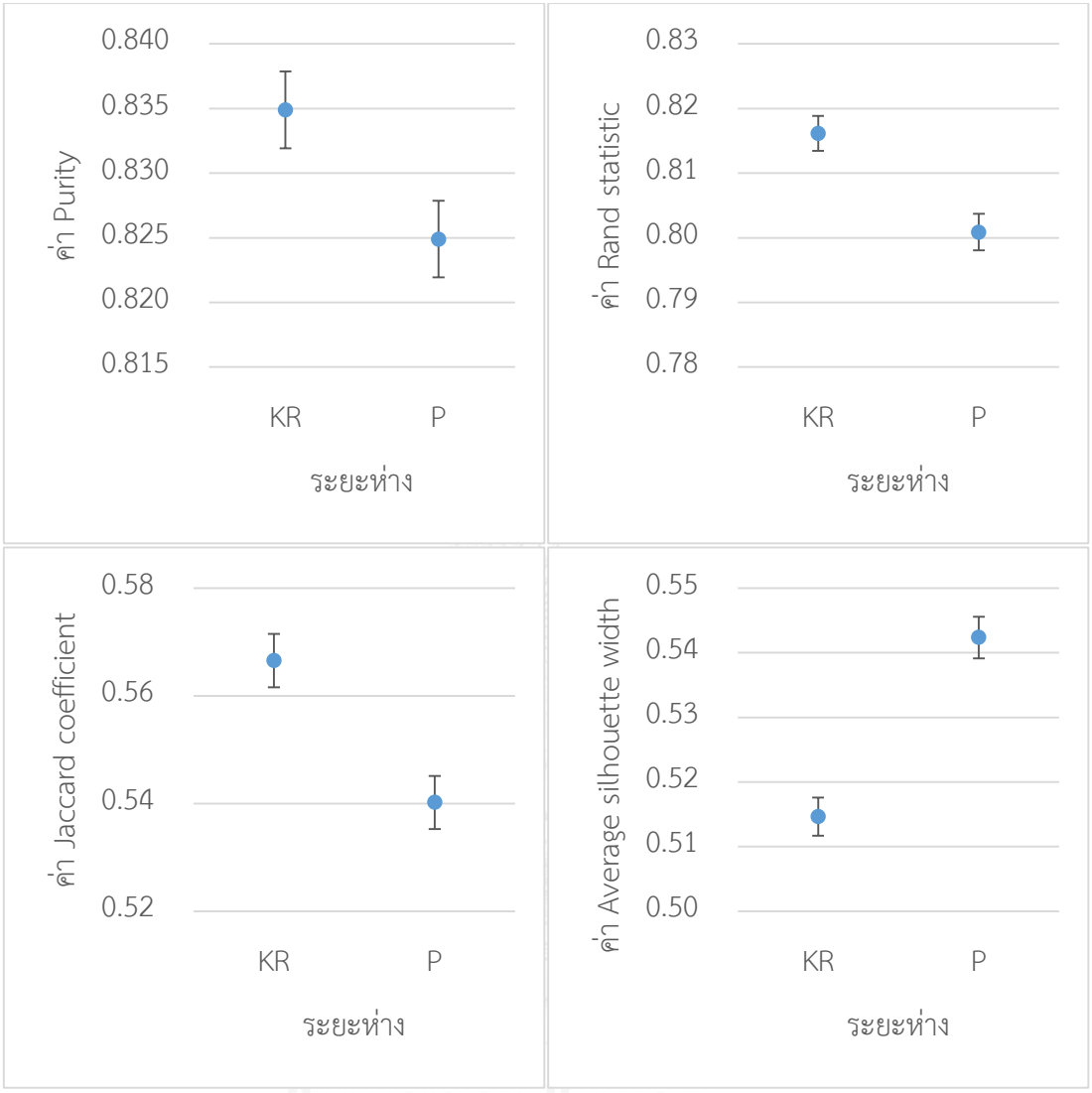
เมื่อพิจารณากราฟช่วงความเชื่อมั่น 95% ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR และระยะห่างของ P สำหรับข้อมูลรูปแบบที่ IV พบว่า เมื่อ $K = 3$ การวิเคราะห์กลุ่มด้วยระยะห่างของ KR มีประสิทธิภาพดีกว่าการวิเคราะห์กลุ่มด้วยระยะห่างของ P ทุกกรณี โดยไม่มีส่วนที่ทับซ้อนกันของช่วงความเชื่อมั่น 95% แต่ค่าเฉลี่ยของค่า Average silhouette width จากการวิเคราะห์กลุ่มด้วยระยะห่างของ P มีค่ามากกว่าจากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR ดังรูปที่ 4.25 ถึง 4.28



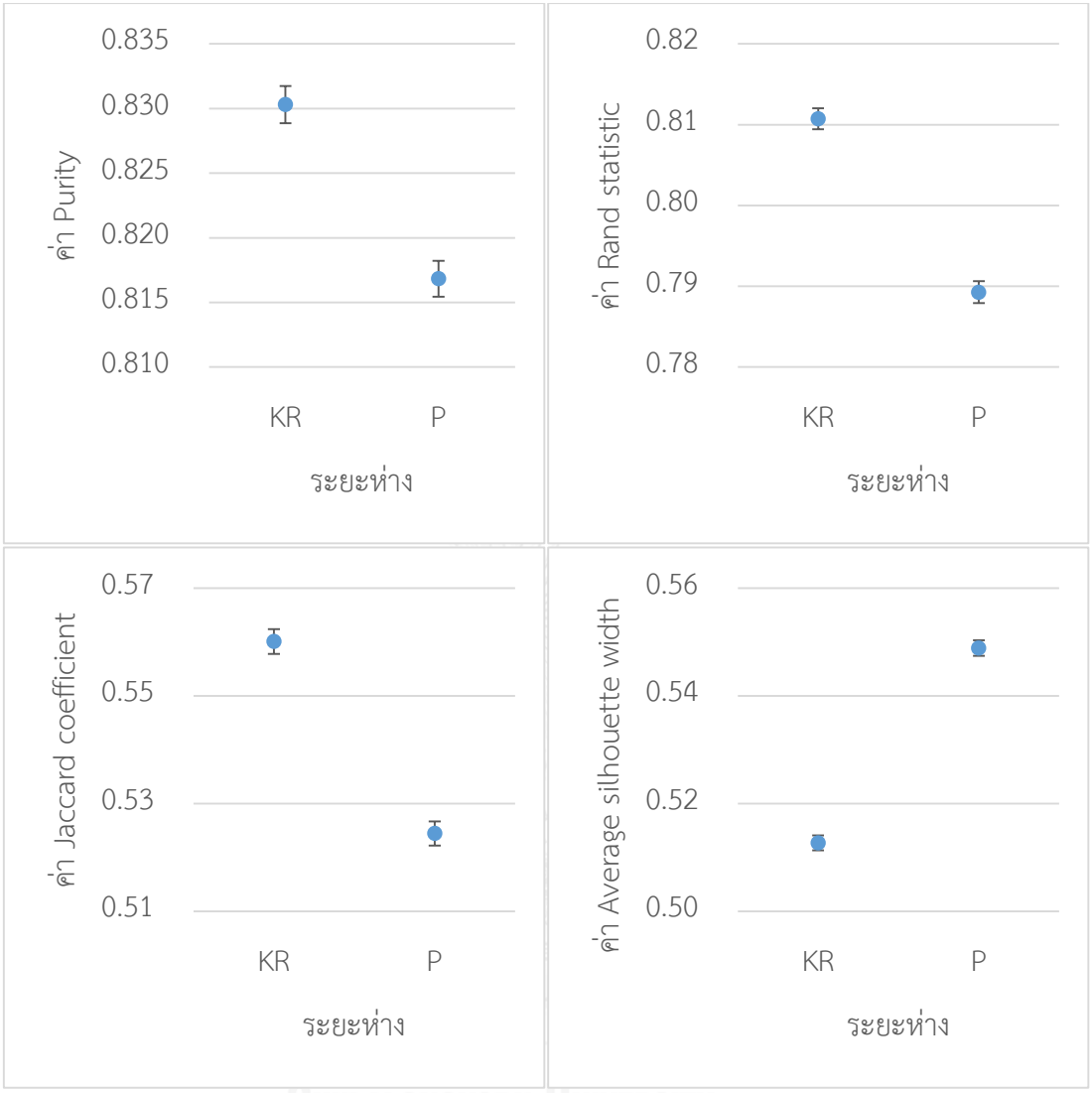
รูปที่ 4.25 กราฟช่วงความเชื่อมั่น 95% ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ IV เมื่อ $K = 3$, $\rho = 0.2$ และ $n = 20$



รูปที่ 4.26 กราฟช่วงความเชื่อมั่น 95% ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ IV เมื่อ $K = 3$, $\rho = 0.2$ และ $n = 100$

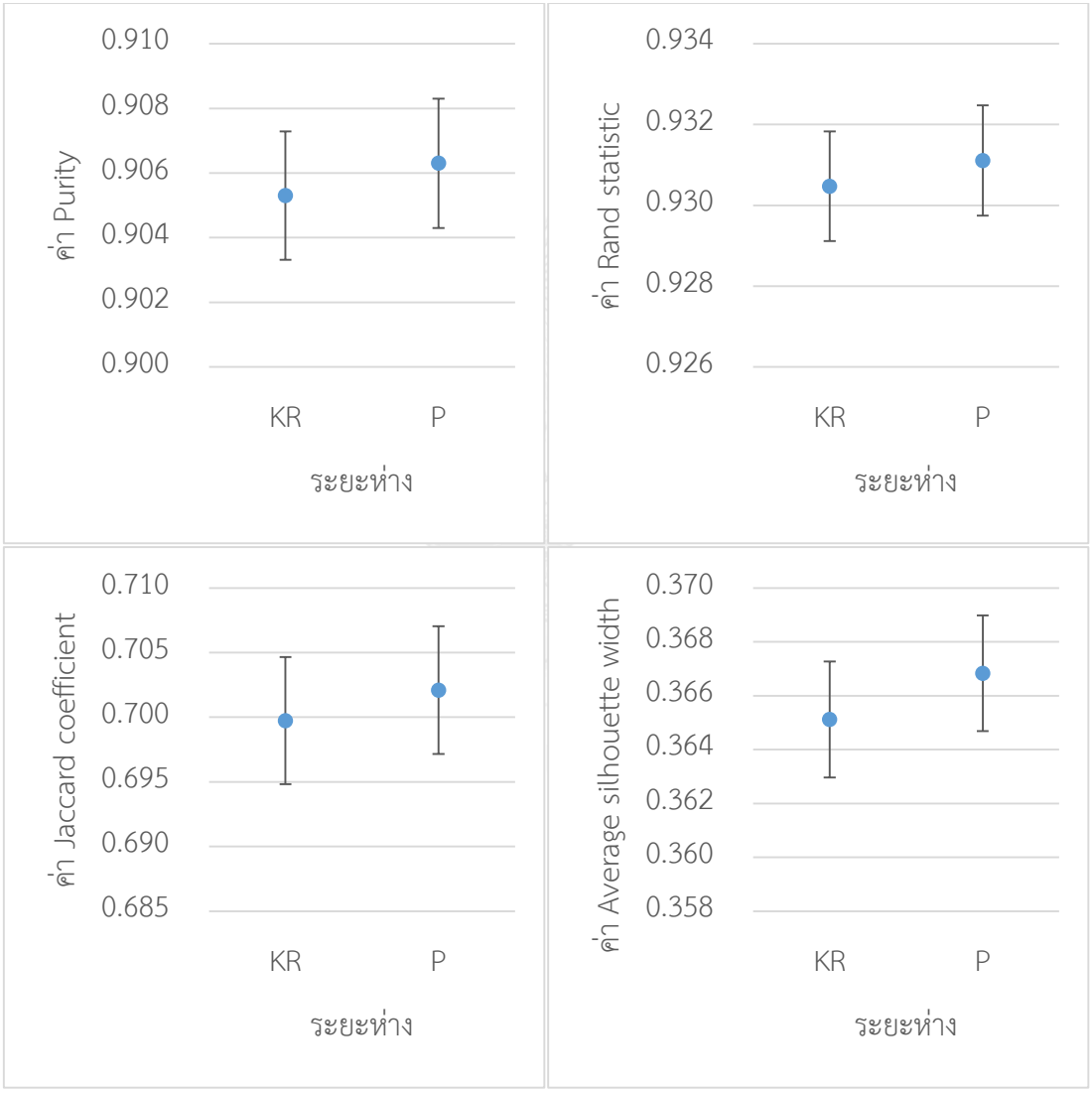


รูปที่ 4.27 กราฟช่วงความเชื่อมั่น 95% ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ IV เมื่อ $K = 3$, $\rho = 0.8$ และ $n = 20$

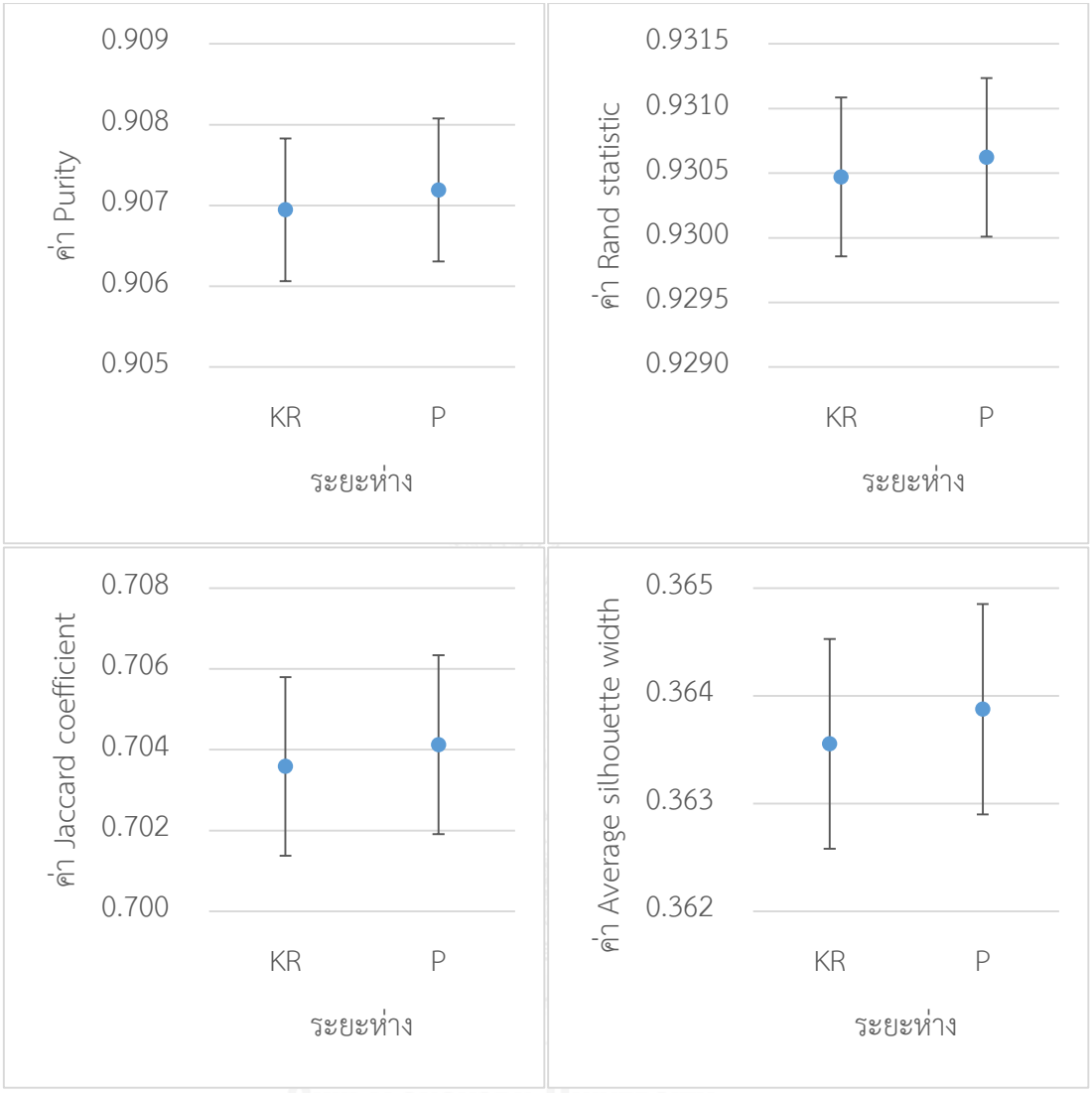


รูปที่ 4.28 กราฟช่วงความเชื่อมั่น 95% ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ IV เมื่อ $K = 3$, $\rho = 0.8$ และ $n = 100$

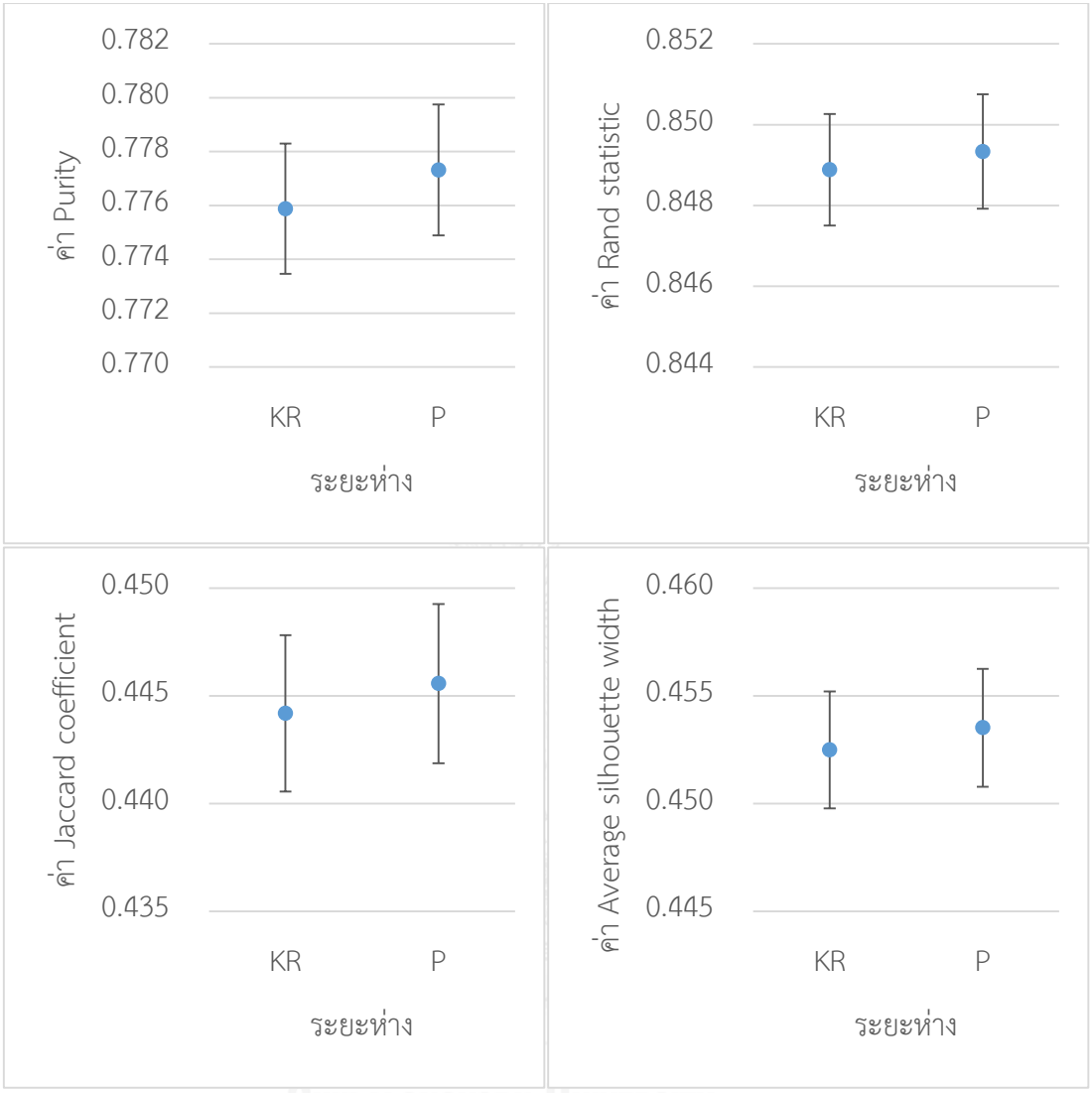
เมื่อ K = 5 แม้ว่าการวิเคราะห์กลุ่มด้วยระยะห่างของ P มีประสิทธิภาพดีกว่าการวิเคราะห์กลุ่มด้วยระยะห่างของ KR โดยเฉลี่ย แต่พบว่าการวิเคราะห์กลุ่มด้วยระยะห่างทั้ง 2 วิธีนี้ มีช่วงความเชื่อมั่น 95% ที่ทับซ้อนกันอยู่ในทุกกรณี นอกจากนี้ค่าเฉลี่ยของค่า Average silhouette width จากการวิเคราะห์กลุ่มด้วยระยะห่างของ P มีค่ามากกว่าจากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR และมีช่วงความเชื่อมั่น 95% ที่ทับซ้อนกันอยู่เช่นเดียวกัน ดังรูปที่ 4.29 และ 4.32



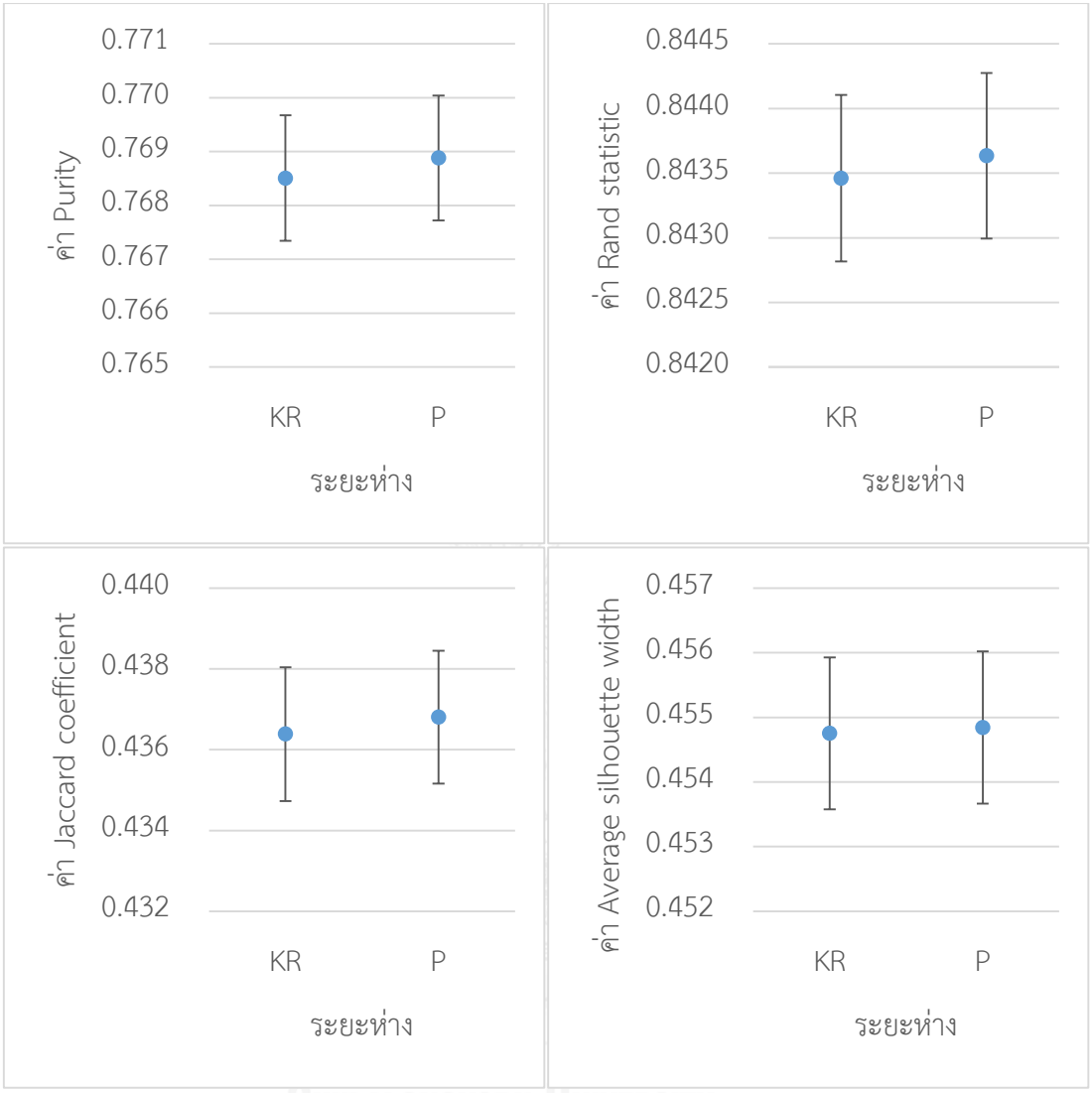
รูปที่ 4.29 กราฟช่วงความเชื่อมั่น 95% ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ IV เมื่อ K = 5, ρ = 0.2 และ n = 20



รูปที่ 4.30 กราฟช่วงความเชื่อมั่น 95% ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ IV เมื่อ $K = 5$, $\rho = 0.2$ และ $n = 100$



รูปที่ 4.31 กราฟช่วงความเชื่อมั่น 95% ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ IV เมื่อ $K = 5$, $\rho = 0.8$ และ $n = 20$



รูปที่ 4.32 กราฟช่วงความเชื่อมั่น 95% ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ IV เมื่อ $K = 5$, $\rho = 0.8$ และ $n = 100$

เนื่องจากการพิจารณากราฟช่วงความเชื่อมั่น 95% พบว่าค่าเฉลี่ยของค่า Purity ค่า Rand statistic ค่า Jaccard coefficient และค่า Average silhouette width มีค่าใกล้เคียงกันในบางกรณี ทั้งยังมีส่วนของช่วงความเชื่อมั่น 95% ทับซ้อนกัน จึงไม่สามารถสรุปได้ว่าการวิเคราะห์กลุ่มด้วยระยะห่างแบบใดมีประสิทธิภาพดีที่สุด ผู้วิจัยจึงทดสอบความแตกต่างระหว่างค่าเฉลี่ยของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแต่ละคู่ ด้วยสถิติทดสอบ t และแบ่งการพิจารณาข้อมูลรูปแบบที่ IV เป็นกรณีใหญ่ 2 กรณีตามจำนวนกลุ่มข้อมูล คือ กรณีที่ $K = 3$ และ $K = 5$

4.1.4.1 ผลทดสอบความแตกต่างการวิเคราะห์กลุ่มข้อมูลรูปแบบที่ IV เมื่อ K = 3

เมื่อ K = 3 ข้อมูลยังถูกแบ่งออกเป็นกรณีต่าง ๆ ตามค่าสัมประสิทธิ์สหสัมพันธ์และขนาดข้อมูลต่อกลุ่ม จึงได้ข้อมูลที่แตกต่างกันดังต่อไปนี้

ข้อมูลรูปแบบที่ IV เมื่อ K = 3, ρ = 0.2 และ n = 20

ข้อมูลรูปแบบที่ IV เมื่อ K = 3, ρ = 0.2 และ n = 100

ข้อมูลรูปแบบที่ IV เมื่อ K = 3, ρ = 0.8 และ n = 20

ข้อมูลรูปแบบที่ IV เมื่อ K = 3, ρ = 0.8 และ n = 100

ทำการศึกษาประสิทธิภาพการวิเคราะห์กลุ่มข้อมูลต่าง ๆ โดยทดสอบความแตกต่างของค่าเฉลี่ยของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างระหว่างระยะห่างของ KR และระยะห่างของ P ด้วยค่าสถิติทดสอบ t ดังตารางที่ 4.45 ถึง 4.48 พบว่าในทุกกรณี ค่า Purity ค่า Rand statistic ค่า Jaccard coefficient และค่า Average silhouette width จากการวิเคราะห์กลุ่ม มี Sig. น้อยกว่า 0.05 จึงปฏิเสธสมมติฐานหลัก ดังนั้น การวิเคราะห์กลุ่มด้วยระยะห่างของ KR มีประสิทธิภาพและมีค่า Average silhouette width แตกต่างจากการวิเคราะห์กลุ่มด้วยระยะห่างของ P ที่ระดับนัยสำคัญ 0.05

ตารางที่ 4.45 ผลทดสอบความแตกต่างระหว่างค่าเฉลี่ยของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR และระยะห่างของ P สำหรับข้อมูลรูปแบบที่ IV เมื่อ K = 3, ρ = 0.2 และ n = 20 ด้วยสถิติทดสอบ t

ค่าที่วัด	t	df	Sig. (2-tailed)	ผลต่างค่าเฉลี่ย	ผลต่าง Std. Error
ค่า Purity	9.2977	1982.2081	**0.0000**	0.0137	0.0015
ค่า Rand statistic	9.6811	1987.3397	**0.0000**	0.0164	0.0017
ค่า Jaccard coefficient	9.7596	1998.0000	**0.0000**	0.0401	0.0041
ค่า Average silhouette width	-8.2160	1982.5158	**0.0000**	-0.0151	0.0018

ตารางที่ 4.46 ผลทดสอบความแตกต่างระหว่างค่าเฉลี่ยของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR และระยะห่างของ P สำหรับข้อมูลรูปแบบที่ IV เมื่อ K = 3, ρ = 0.2 และ n = 100 ด้วยสถิติทดสอบ t

ค่าที่วัด	t	df	Sig. (2-tailed)	ผลต่างค่าเฉลี่ย	ผลต่าง Std. Error
ค่า Purity	19.5750	1985.2423	**0.0000**	0.0136	0.0007
ค่า Rand statistic	20.5221	1998.0000	**0.0000**	0.0168	0.0008
ค่า Jaccard coefficient	20.5372	1998.0000	**0.0000**	0.0403	0.0020
ค่า Average silhouette width	-19.5783	1984.0898	**0.0000**	-0.0163	0.0008

ตารางที่ 4.47 ผลทดสอบความแตกต่างระหว่างค่าเฉลี่ยของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR และระยะห่างของ P สำหรับข้อมูลรูปแบบที่ IV เมื่อ K = 3, ρ = 0.8 และ n = 20 ด้วยสถิติทดสอบ t

ค่าที่วัด	t	df	Sig. (2-tailed)	ผลต่างค่าเฉลี่ย	ผลต่าง Std. Error
ค่า Purity	4.6691	1998.0000	**0.0000**	0.0100	0.0021
ค่า Rand statistic	7.6448	1998.0000	**0.0000**	0.0153	0.0020
ค่า Jaccard coefficient	7.3746	1998.0000	**0.0000**	0.0263	0.0036
ค่า Average silhouette width	-12.4466	1983.3506	**0.0000**	-0.0277	0.0022

ตารางที่ 4.48 ผลทดสอบความแตกต่างระหว่างค่าเฉลี่ยของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR และระยะห่างของ P สำหรับข้อมูลรูปแบบที่ IV เมื่อ K = 3, ρ = 0.8 และ n = 100 ด้วยสถิติทดสอบ t

ค่าที่วัด	t	df	Sig. (2-tailed)	ผลต่างค่าเฉลี่ย	ผลต่าง Std. Error
ค่า Purity	9.2977	1982.2081	0.0000	0.0137	0.0015
ค่า Rand statistic	9.6811	1987.3397	**0.0000**	0.0164	0.0017
ค่า Jaccard coefficient	9.7596	1998.0000	**0.0000**	0.0401	0.0041
ค่า Average silhouette width	-8.2160	1982.5158	**0.0000**	-0.0151	0.0018

4.1.4.2 ผลทดสอบความแตกต่างการวิเคราะห์กลุ่มข้อมูลรูปแบบที่ IV เมื่อ $K = 5$

เมื่อ $K = 5$ ทำการทดสอบความแตกต่างของประสิทธิภาพการวิเคราะห์กลุ่มด้วยระยะห่างของ KR และระยะห่างของ P ในกรณีที่ข้อมูลแตกต่างกันดังต่อไปนี้

ข้อมูลรูปแบบที่ IV เมื่อ $K = 5$, $\rho = 0.2$ และ $n = 20$

ข้อมูลรูปแบบที่ IV เมื่อ $K = 5$, $\rho = 0.2$ และ $n = 100$

ข้อมูลรูปแบบที่ IV เมื่อ $K = 5$, $\rho = 0.8$ และ $n = 20$

ข้อมูลรูปแบบที่ IV เมื่อ $K = 5$, $\rho = 0.8$ และ $n = 100$

แม้ว่าการวิเคราะห์กลุ่มด้วยระยะห่างของ P มีประสิทธิภาพดีกว่าการวิเคราะห์กลุ่มด้วยระยะห่างของ KR โดยเฉลี่ย ดังตารางที่ 4.44 แต่เมื่อทำการศึกษาประสิทธิภาพการวิเคราะห์กลุ่มข้อมูลต่าง ๆ โดยทดสอบความแตกต่างของค่าเฉลี่ยของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างระหว่างระยะห่างของ KR และระยะห่างของ P ด้วยค่าสถิติทดสอบ t ดังตารางที่ 4.49 ถึง 4.52 พบว่าในทุกกรณี ค่า Purity ค่า Rand statistic ค่า Jaccard coefficient และค่า Average silhouette width จากการวิเคราะห์กลุ่ม มี Sig. มากกว่า 0.05 จึงไม่สามารถปฏิเสธสมมติฐานหลัก ดังนั้นการวิเคราะห์กลุ่มด้วยระยะห่างของ KR มีประสิทธิภาพและมีค่า Average silhouette width ไม่แตกต่างจากการวิเคราะห์กลุ่มด้วยระยะห่างของ P

ตารางที่ 4.49 ผลทดสอบความแตกต่างระหว่างค่าเฉลี่ยของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR และระยะห่างของ P สำหรับข้อมูลรูปแบบที่ IV เมื่อ $K = 5$, $\rho = 0.2$ และ $n = 20$ ด้วยสถิติทดสอบ t

ค่าที่วัด	t	df	Sig. (2-tailed)	ผลต่างค่าเฉลี่ย	ผลต่าง Std. Error
ค่า Purity	-0.6955	1998.0000	0.4868	-0.0010	0.0014
ค่า Rand statistic	-0.6504	1998.0000	0.5155	-0.0006	0.0010
ค่า Jaccard coefficient	-0.6646	1998.0000	0.5064	-0.0024	0.0036
ค่า Average silhouette width	-1.1086	1998.0000	0.2677	-0.0017	0.0016

ตารางที่ 4.50 ผลทดสอบความแตกต่างระหว่างค่าเฉลี่ยของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR และระยะห่างของ P สำหรับข้อมูลรูปแบบที่ IV เมื่อ K = 5, ρ = 0.2 และ n = 100 ด้วยสถิติทดสอบ t

ค่าที่วัด	t	df	Sig. (2-tailed)	ผลต่างค่าเฉลี่ย	ผลต่าง Std. Error
ค่า Purity	-0.3833	1998.0000	0.7015	-0.0002	0.0006
ค่า Rand statistic	-0.3414	1998.0000	0.7328	-0.0002	0.0004
ค่า Jaccard coefficient	-0.3366	1998.0000	0.7365	-0.0005	0.0016
ค่า Average silhouette width	-0.4567	1998.0000	0.6479	-0.0003	0.0007

ตารางที่ 4.51 ผลทดสอบความแตกต่างระหว่างค่าเฉลี่ยของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR และระยะห่างของ P สำหรับข้อมูลรูปแบบที่ IV เมื่อ K = 5, ρ = 0.8 และ n = 20 ด้วยสถิติทดสอบ t

ค่าที่วัด	t	df	Sig. (2-tailed)	ผลต่างค่าเฉลี่ย	ผลต่าง Std. Error
ค่า Purity	-0.3833	1998.0000	0.7015	-0.0002	0.0006
ค่า Rand statistic	-0.3414	1998.0000	0.7328	-0.0002	0.0004
ค่า Jaccard coefficient	-0.3366	1998.0000	0.7365	-0.0005	0.0016
ค่า Average silhouette width	-0.4567	1998.0000	0.6479	-0.0003	0.0007

ตารางที่ 4.52 ผลทดสอบความแตกต่างระหว่างค่าเฉลี่ยของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR และระยะห่างของ P สำหรับข้อมูลรูปแบบที่ IV เมื่อ K = 5, ρ = 0.8 และ n = 100 ด้วยสถิติทดสอบ t

ค่าที่วัด	t	df	Sig. (2-tailed)	ผลต่างค่าเฉลี่ย	ผลต่าง Std. Error
ค่า Purity	-0.3833	1998.0000	0.7015	-0.0002	0.0006
ค่า Rand statistic	-0.3414	1998.0000	0.7328	-0.0002	0.0004
ค่า Jaccard coefficient	-0.3366	1998.0000	0.7365	-0.0005	0.0016
ค่า Average silhouette width	-0.4567	1998.0000	0.6479	-0.0003	0.0007

4.1.5 ผลการวิเคราะห์กลุ่มข้อมูลรูปแบบที่ V

ผู้วิจัยทำการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ 2 วิธี ได้แก่ ระยะห่างของ KR และระยะห่างแบบ KR&N สำหรับข้อมูลรูปแบบที่ V เนื่องจากเป็นข้อมูลที่ประกอบไปด้วยตัวแปรนามบัญญัติเพียงชนิดเดียว จำนวน 3 ตัวแปร และคำนวณค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของ ค่า Purity ค่า Rand statistic ค่า Jaccard coefficient และค่า Average silhouette width ได้ผลดังตารางที่ 4.53

ผลการเปรียบเทียบประสิทธิภาพโดยเฉลี่ยในการวิเคราะห์กลุ่มด้วยมาตรวัดระยะห่างแบบต่าง ๆ พบว่า การวิเคราะห์กลุ่มด้วยระยะห่างของ KR ให้ค่า Purity ค่า Rand statistic และค่า Jaccard coefficient สูงกว่าการวิเคราะห์กลุ่มด้วยระยะห่างแบบ KR&N ทุกกรณี ยกเว้นกรณีที่ K = 5, ρ = 0.8 และ n = 100 นอกจากนี้ เมื่อพิจารณาค่าเฉลี่ยของค่า Average silhouette width จากการวิเคราะห์กลุ่มด้วยระยะห่างทั้ง 2 วิธี พบว่า เมื่อ K = 3 ค่าเฉลี่ยของค่า Average silhouette width จากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR มีค่ามากกว่าจากการวิเคราะห์กลุ่มด้วยระยะห่างแบบ KR&N แต่เมื่อ K = 5 ค่าเฉลี่ยของค่า Average silhouette width จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบ KR&N มีค่ามากกว่าจากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR

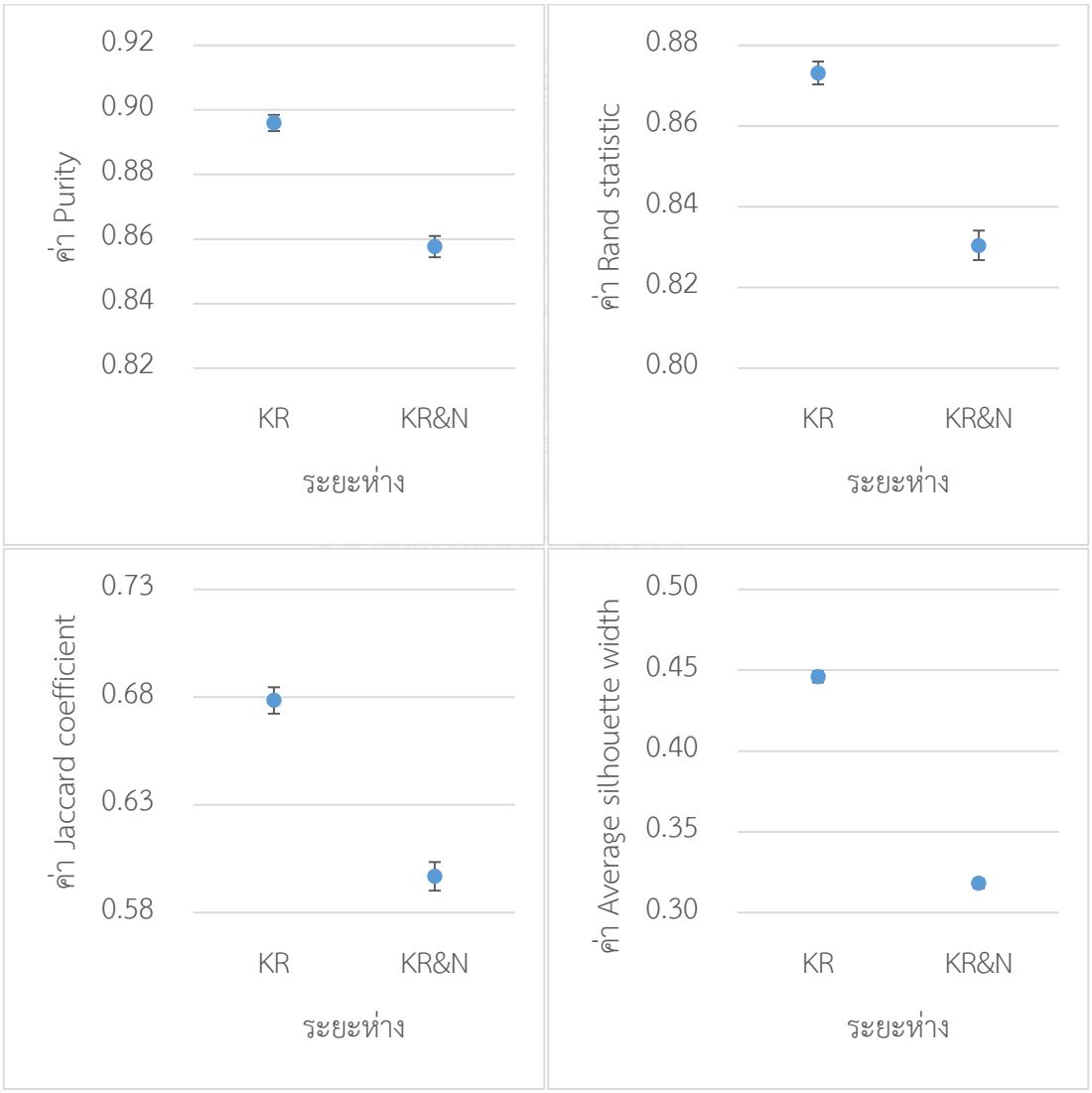
ตารางที่ 4.53 ค่าเฉลี่ย (ส่วนเบี่ยงเบนมาตรฐาน) จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ V

K	ρ	n	ค่าที่วัด	ระยะห่าง	
				KR	KR&N
3	0.2	20	ค่า Purity	**0.8959** (0.0409)	0.8577 (0.0533)
			ค่า Rand statistic	**0.8731** (0.0459)	0.8304 (0.0594)
			ค่า Jaccard coefficient	**0.6785** (0.0989)	0.5968 (0.1081)
			ค่า Average silhouette width	**0.4458** (0.0565)	0.3180 (0.0450)
		100	ค่า Purity	**0.8925** (0.0208)	0.8759 (0.0201)
			ค่า Rand statistic	**0.8681** (0.0236)	0.8500 (0.0218)
			ค่า Jaccard coefficient	**0.6703** (0.0485)	0.6330 (0.0435)
			ค่า Average silhouette width	**0.4494** (0.0247)	0.2954 (0.0200)
	0.8	20	ค่า Purity	**0.8077** (0.0518)	0.7907 (0.0547)
			ค่า Rand statistic	**0.7812** (0.0502)	0.7448 (0.0686)
			ค่า Jaccard coefficient	**0.5042** (0.0843)	0.4598 (0.0908)
			ค่า Average silhouette width	**0.5387** (0.0681)	0.4049 (0.0602)

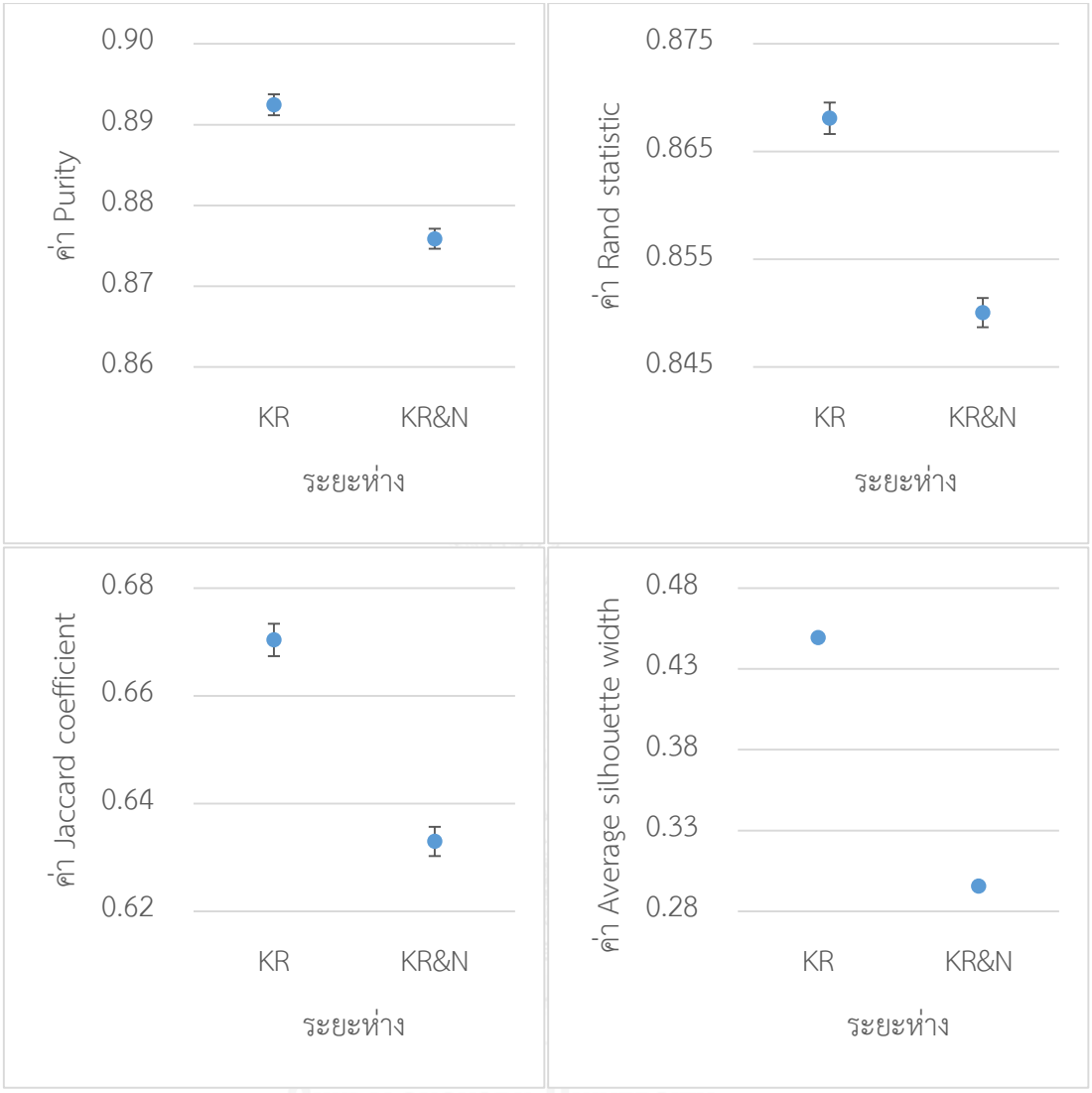
ตารางที่ 4.53 (ต่อ) ค่าเฉลี่ย (ส่วนเบี่ยงเบนมาตรฐาน) จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ V

K	ρ	n	ค่าที่วัด	ระยะห่าง	
				KR	KR&N
3	0.8	100	ค่า Purity	**0.8024** (0.0231)	0.7937 (0.0257)
			ค่า Rand statistic	**0.7730** (0.0226)	0.7527 (0.0476)
			ค่า Jaccard coefficient	**0.4966** (0.0356)	0.4709 (0.0519)
			ค่า Average silhouette width	**0.5472** (0.0297)	0.3629 (0.0339)
5	0.2	20	ค่า Purity	**0.8255** (0.0387)	0.8090 (0.0390)
			ค่า Rand statistic	**0.8761** (0.0247)	0.8670 (0.0237)
			ค่า Jaccard coefficient	**0.5226** (0.0712)	0.4938 (0.0671)
			ค่า Average silhouette width	0.4328 (0.0439)	**0.4372** (0.0432)
		100	ค่า Purity	**0.8147** (0.0198)	0.8088 (0.0182)
			ค่า Rand statistic	**0.8676** (0.0126)	0.8658 (0.0111)
			ค่า Jaccard coefficient	**0.5054** (0.0345)	0.4969 (0.0310)
			ค่า Average silhouette width	0.4348 (0.0198)	**0.4373** (0.0195)
	0.8	20	ค่า Purity	**0.7380** (0.0445)	0.7354 (0.0435)
			ค่า Rand statistic	**0.8277** (0.0238)	0.8260 (0.0231)
			ค่า Jaccard coefficient	**0.3899** (0.0581)	0.3856 (0.0561)
			ค่า Average silhouette width	0.6134 (0.0472)	**0.6139** (0.0467)
		100	ค่า Purity	0.7388 (0.0192)	**0.7402** (0.0185)
			ค่า Rand statistic	0.8256 (0.0103)	**0.8270** (0.0099)
			ค่า Jaccard coefficient	0.3932 (0.0247)	**0.3952** (0.0241)
			ค่า Average silhouette width	0.6176 (0.0206)	**0.6180** (0.0205)

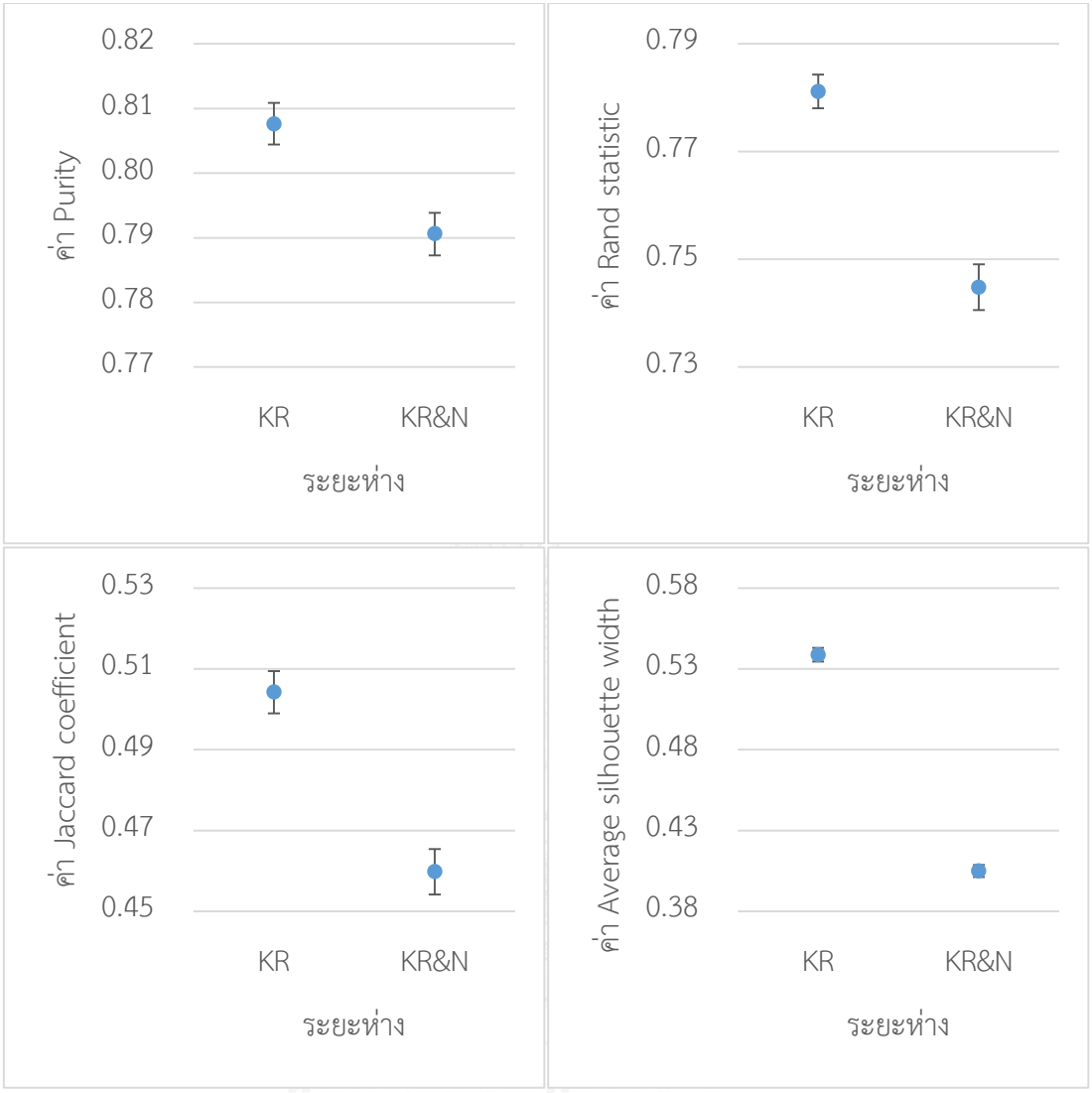
พิจารณากราฟช่วงความเชื่อมั่น 95% ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ V เมื่อ $K = 3$ ค่าเฉลี่ยของค่า Purity ค่า Rand statistic ค่า Jaccard coefficient และค่า Average silhouette width จากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR มากกว่าจากการวิเคราะห์กลุ่มด้วยระยะห่างแบบ KR&N โดยไม่มีส่วนที่ทับซ้อนกันของช่วงความเชื่อมั่น 95% ในทุกกรณี ดังนั้นโดยเฉลี่ยการวิเคราะห์กลุ่มด้วยระยะห่างของ KR มีประสิทธิภาพดีกว่าการวิเคราะห์กลุ่มด้วยระยะห่างแบบ KR&N ทุกกรณี และค่าเฉลี่ยของค่า Average silhouette width จากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR มีค่ามากกว่าจากการวิเคราะห์กลุ่มด้วยระยะห่างแบบ KR&N ในทุกกรณีเช่นเดียวกัน ดังรูปที่ 4.33 ถึง 4.36



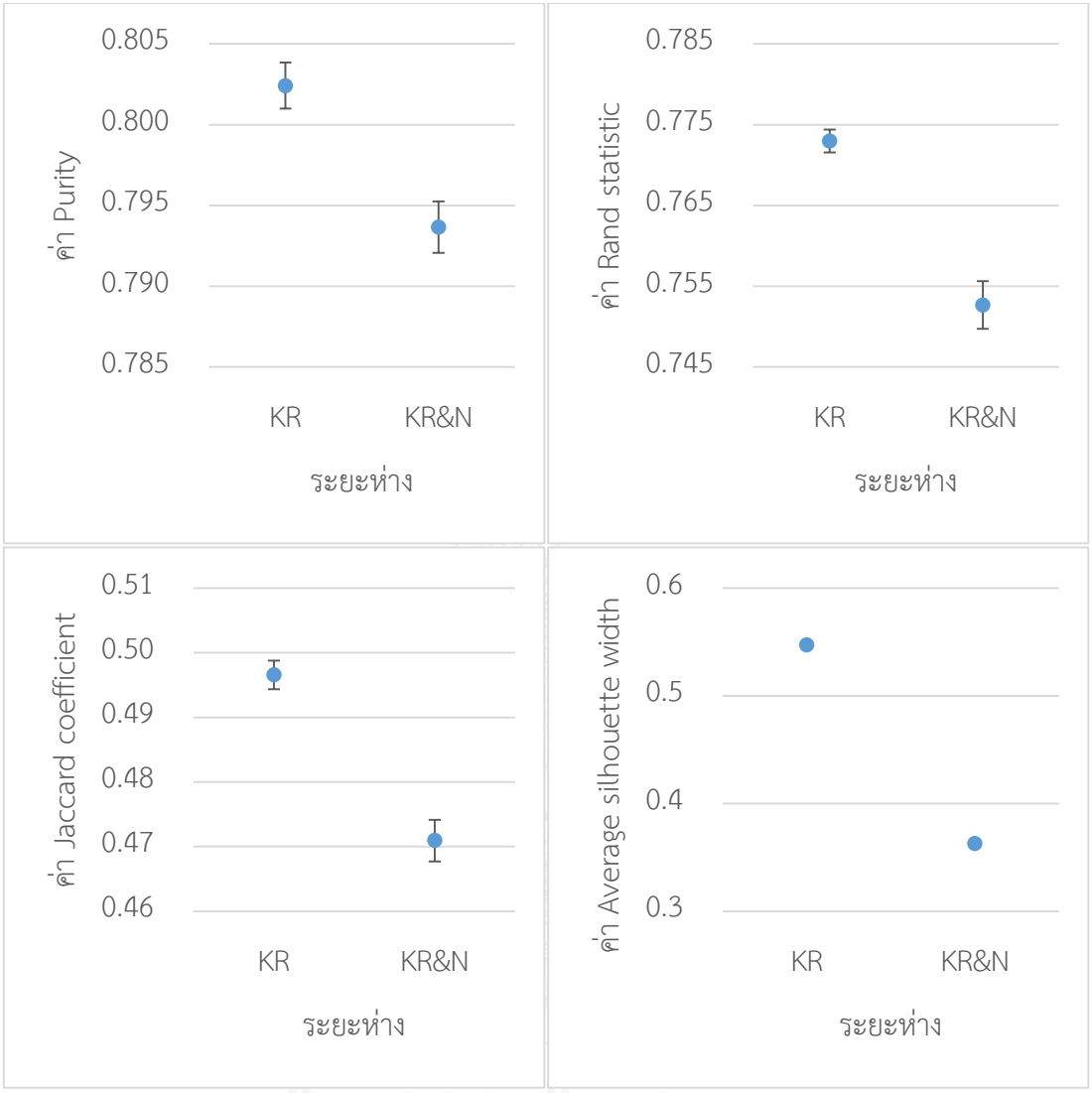
รูปที่ 4.33 กราฟช่วงความเชื่อมั่น 95% ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ V เมื่อ $K = 3$, $\rho = 0.2$ และ $n = 20$



รูปที่ 4.34 กราฟช่วงความเชื่อมั่น 95% ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ V เมื่อ $K = 3$, $\rho = 0.2$ และ $n = 100$

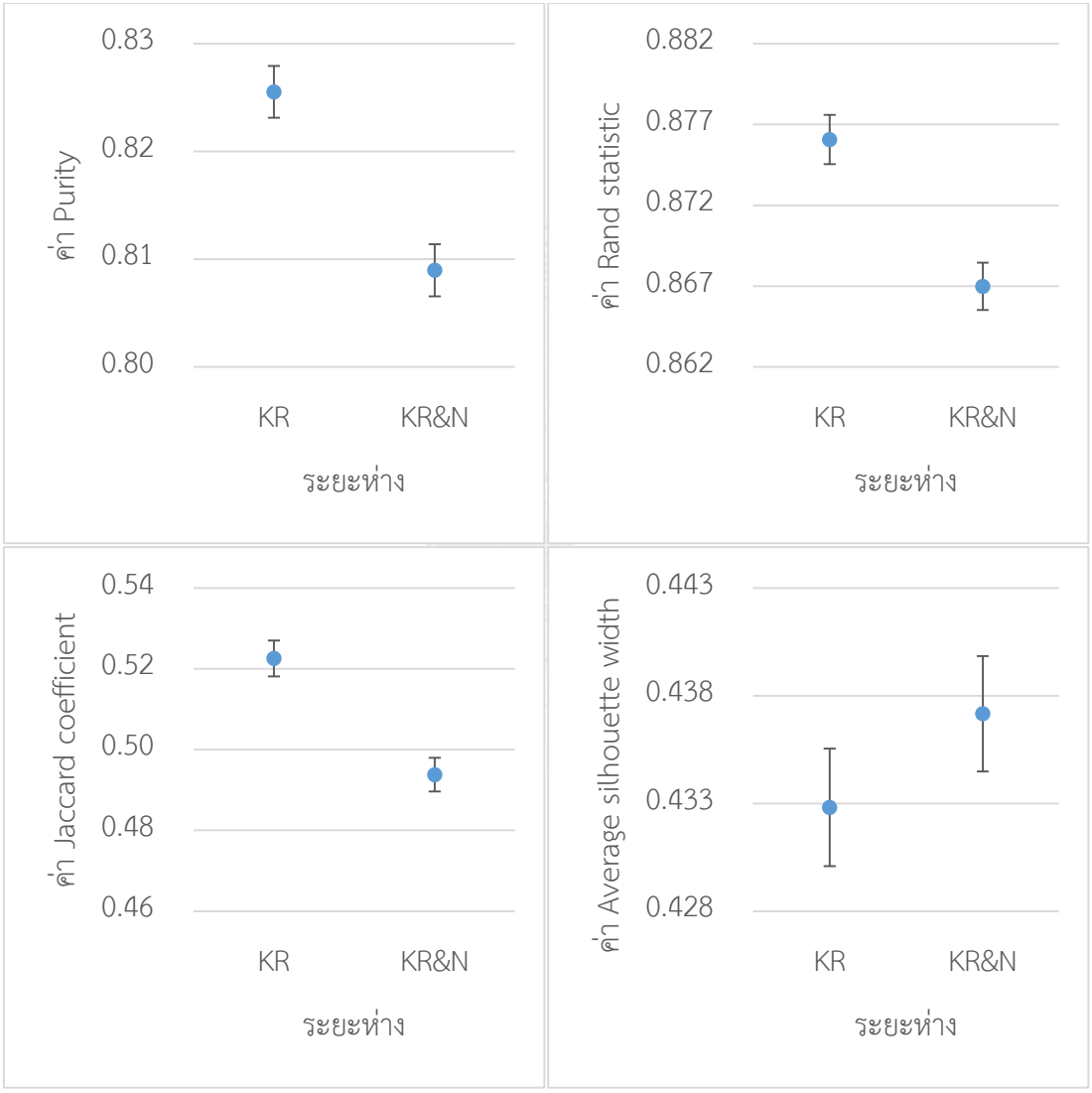


รูปที่ 4.35 กราฟช่วงความเชื่อมั่น 95% ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ V เมื่อ $K = 3$, $\rho = 0.8$ และ $n = 20$

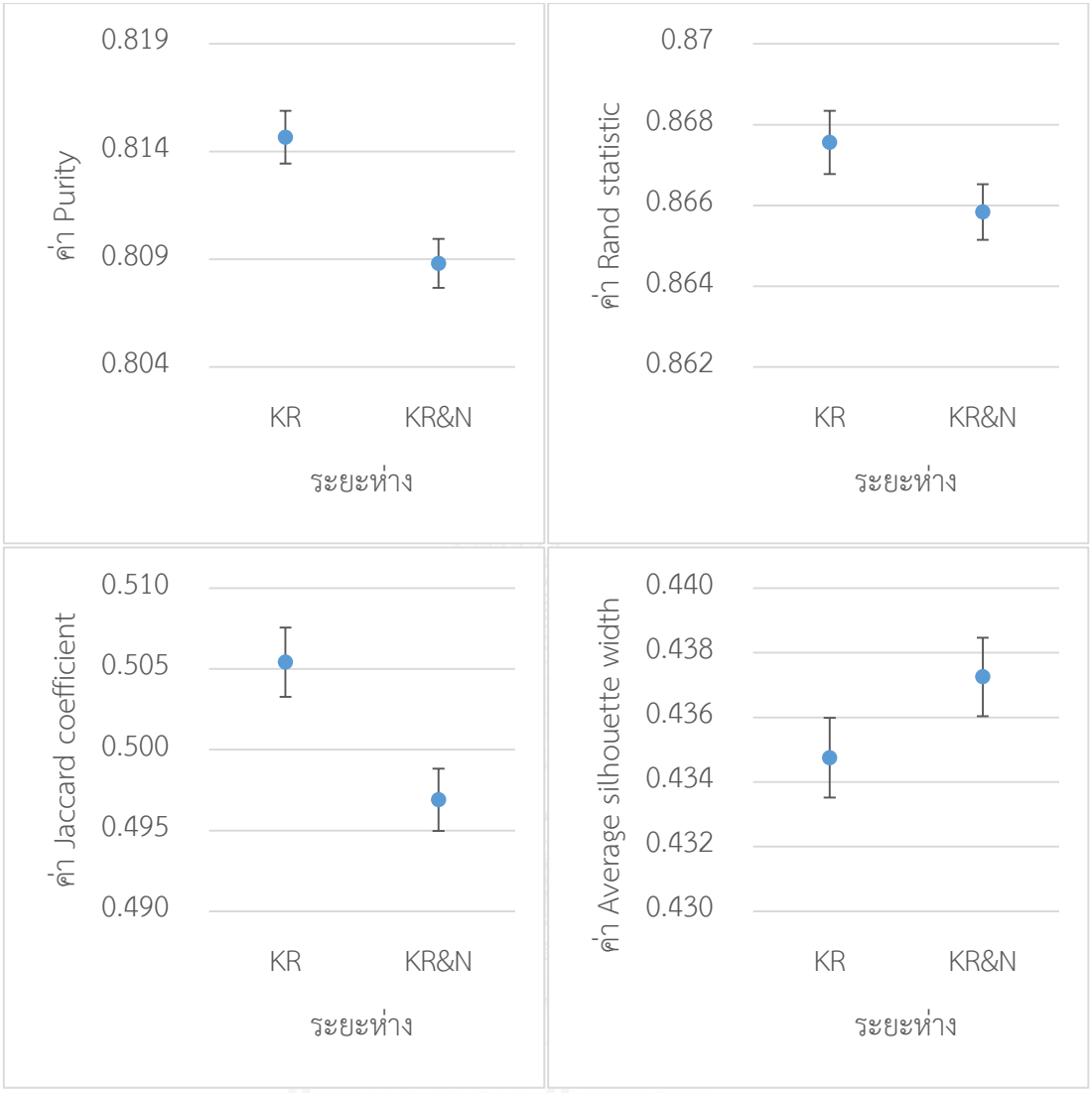


รูปที่ 4.36 กราฟช่วงความเชื่อมั่น 95% ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ V เมื่อ $K = 3$, $\rho = 0.8$ และ $n = 100$

เมื่อ K = 5 และ ρ = 0.2 การวิเคราะห์กลุ่มด้วยระยะห่างของ KR ยังคงมีประสิทธิภาพดีกว่าการวิเคราะห์กลุ่มด้วยระยะห่างแบบ KR&N และไม่มีส่วนที่ช่วงความเชื่อมั่น 95% ทับซ้อนกัน แต่ค่าเฉลี่ยของค่า Average silhouette width จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบ KR&N มีค่ามากกว่าจากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR ดังรูปที่ 4.37 และ 4.38



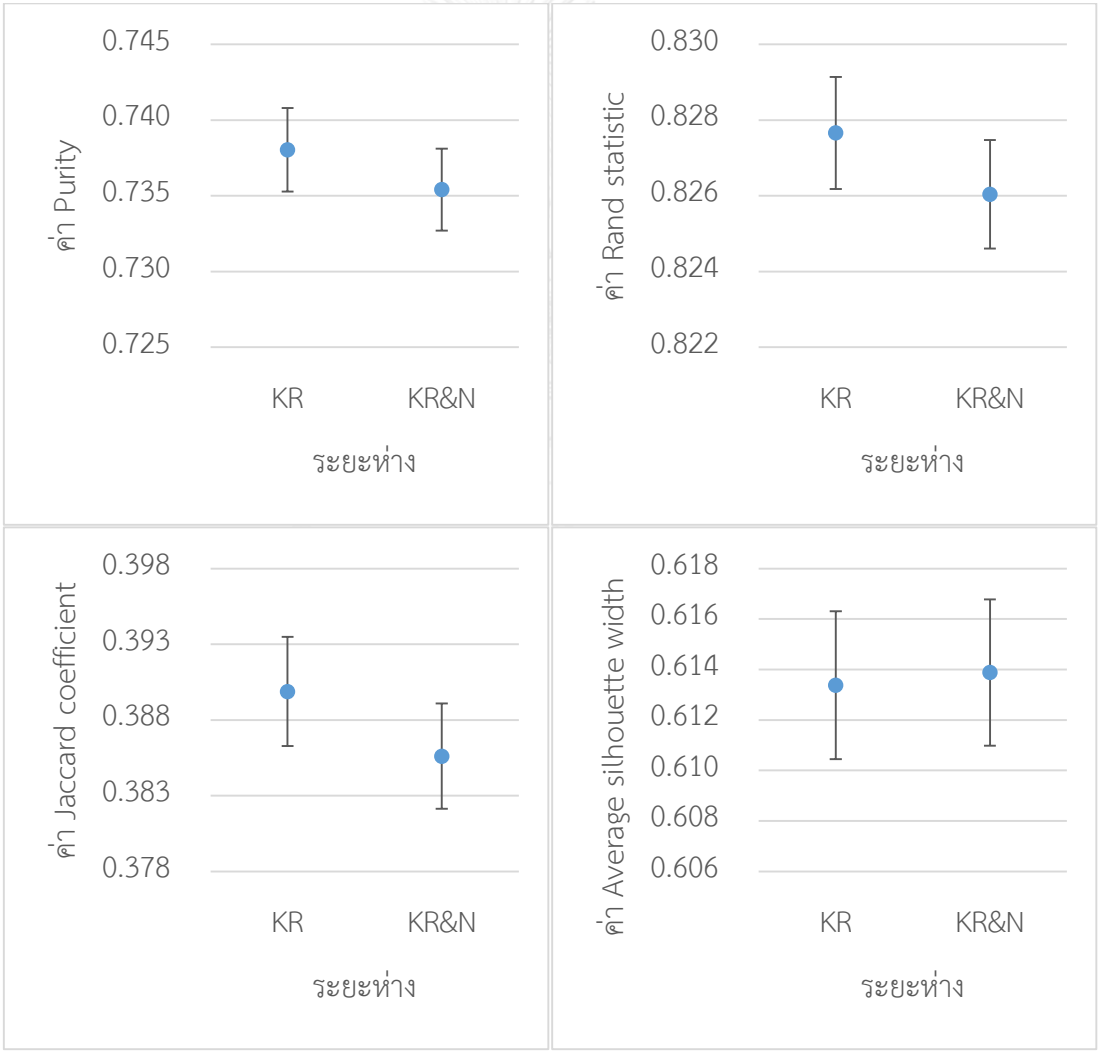
รูปที่ 4.37 กราฟช่วงความเชื่อมั่น 95% ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ V เมื่อ K = 5, ρ = 0.2 และ n = 20



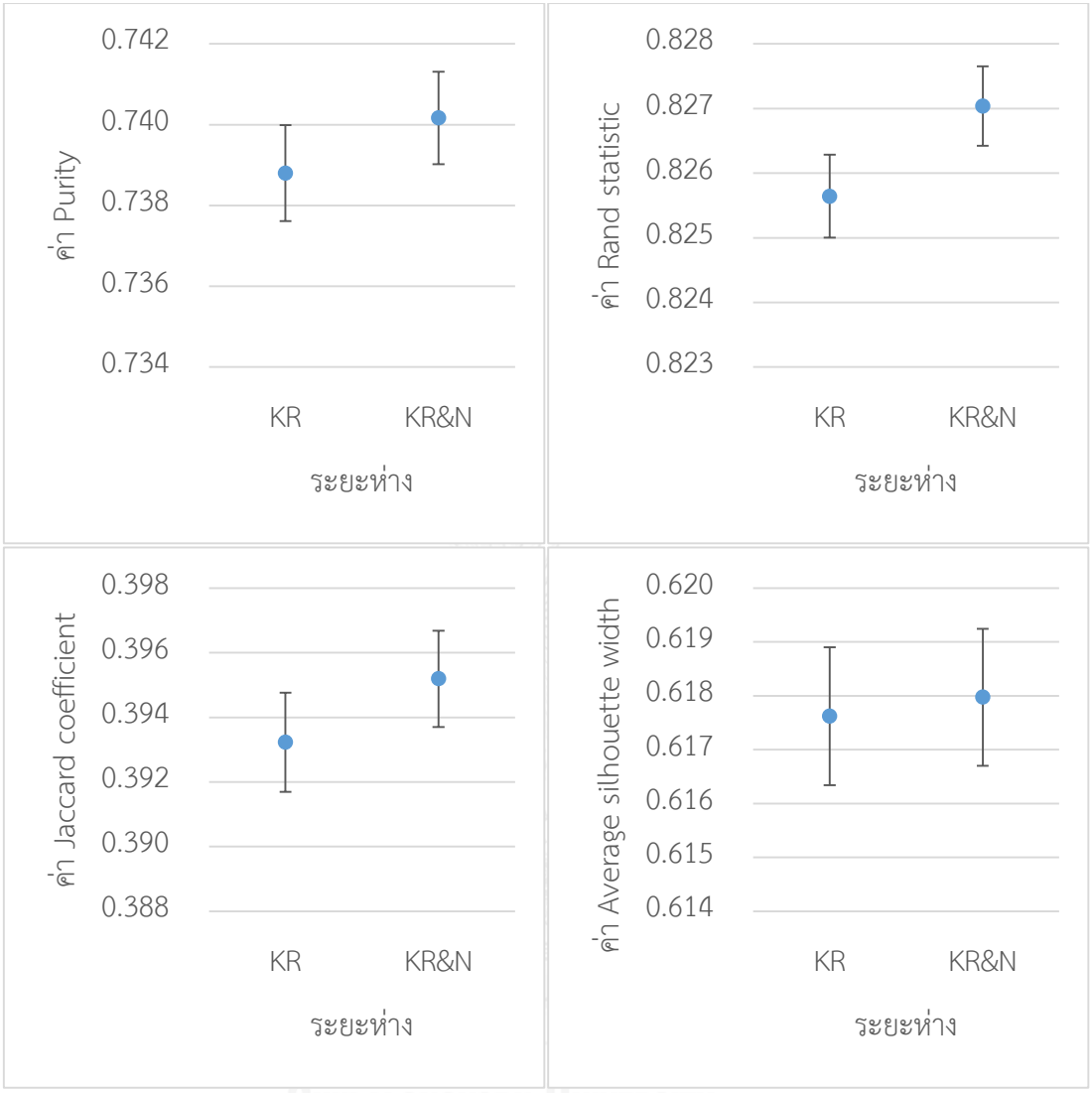
รูปที่ 4.38 กราฟช่วงความเชื่อมั่น 95% ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ V เมื่อ $K = 5$, $\rho = 0.2$ และ $n = 100$

เมื่อ $K = 5$ $\rho = 0.8$ และ $n = 20$ ค่าเฉลี่ยของค่า Purity ค่า Rand statistic และค่า Jaccard coefficient จากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR มากกว่าจากการวิเคราะห์กลุ่มด้วยระยะห่างแบบ KR&N โดยมีส่วนที่ทับซ้อนกันของช่วงความเชื่อมั่น 95% ดังรูปที่ 4.39 แต่เมื่อจำนวนข้อมูลต่อกลุ่มเพิ่มขึ้นเป็น $n = 100$ ค่าเฉลี่ยของค่า Purity ค่า Rand statistic และค่า Jaccard coefficient จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบ KR&N มากกว่าจากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR แต่ยังคงมีส่วนที่ทับซ้อนกันของช่วงความเชื่อมั่น 95% ดังรูปที่ 4.40

แม้ว่าประสิทธิภาพในการวิเคราะห์กลุ่มของกรณีที่ $n = 20$ กับ $n = 100$ จะแตกต่างกัน แต่ค่าเฉลี่ยของค่า Average silhouette width จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบ KR&N กับระยะห่างของ KR มีค่าใกล้เคียงกัน ทั้งสองกรณี



รูปที่ 4.39 กราฟช่วงความเชื่อมั่น 95% ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ V เมื่อ $K = 5$, $\rho = 0.8$ และ $n = 20$



รูปที่ 4.40 กราฟช่วงความเชื่อมั่น 95% ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ V เมื่อ $K = 5$, $\rho = 0.8$ และ $n = 100$

เนื่องจากการพิจารณากราฟช่วงความเชื่อมั่น 95% พบว่าค่าเฉลี่ยของค่า Purity ค่า Rand statistic ค่า Jaccard coefficient และค่า Average silhouette width มีค่าใกล้เคียงกันในบางกรณี ทั้งยังมีส่วนของช่วงความเชื่อมั่น 95% ทับซ้อนกัน จึงไม่สามารถสรุปได้ว่าการวิเคราะห์กลุ่มด้วยระยะห่างแบบใดมีประสิทธิภาพดีที่สุด ผู้วิจัยจึงทดสอบความแตกต่างระหว่างค่าเฉลี่ยของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแต่ละคู่ ด้วยสถิติทดสอบ t และแบ่งการพิจารณาข้อมูลรูปแบบที่ V เป็นกรณีใหญ่ 2 กรณีตามจำนวนกลุ่มข้อมูล คือ กรณีที่ K = 3 และ K = 5

4.1.5.1 ผลทดสอบความแตกต่างการวิเคราะห์กลุ่มข้อมูลรูปแบบที่ V เมื่อ K = 3

ทดสอบความแตกต่างของค่าเฉลี่ยของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างระหว่างระยะห่างของ KR และระยะห่างแบบ KR&N ด้วยค่าสถิติทดสอบ t สำหรับข้อมูลต่อไปนี้

ข้อมูลรูปแบบที่ V เมื่อ K = 3, ρ = 0.2 และ n = 20

ข้อมูลรูปแบบที่ V เมื่อ K = 3, ρ = 0.2 และ n = 100

ข้อมูลรูปแบบที่ V เมื่อ K = 3, ρ = 0.8 และ n = 20

ข้อมูลรูปแบบที่ V เมื่อ K = 3, ρ = 0.8 และ n = 100

ได้ผลดังตารางที่ 4.54 ถึง 4.57 พบว่าในทุกกรณี ค่า Purity ค่า Rand statistic ค่า Jaccard coefficient และค่า Average silhouette width จากการวิเคราะห์กลุ่ม มี Sig. น้อยกว่า 0.05 จึงปฏิเสธสมมติฐานหลัก ดังนั้นการวิเคราะห์กลุ่มด้วยระยะห่างของ KR มีประสิทธิภาพและมีค่า Average silhouette width แตกต่างจากการวิเคราะห์กลุ่มด้วยระยะห่างแบบ KR&N ที่ระดับนัยสำคัญ 0.05

ตารางที่ 4.54 ผลทดสอบความแตกต่างระหว่างค่าเฉลี่ยของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR และระยะห่างแบบ KR&N สำหรับข้อมูลรูปแบบที่ V เมื่อ K = 3, ρ = 0.2 และ n = 20 ด้วยสถิติทดสอบ t

ค่าที่วัด	t	df	Sig. (2-tailed)	ผลต่างค่าเฉลี่ย	ผลต่าง Std. Error
ค่า Purity	17.9957	1873.6994	**0.0000**	0.0382	0.0021
ค่า Rand statistic	17.9720	1878.5027	**0.0000**	0.0427	0.0024
ค่า Jaccard coefficient	17.6407	1982.4633	**0.0000**	0.0817	0.0046
ค่า Average silhouette width	55.9321	1903.1309	**0.0000**	0.1278	0.0023

ตารางที่ 4.55 ผลทดสอบความแตกต่างระหว่างค่าเฉลี่ยของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR และระยะห่างแบบ KR&N สำหรับข้อมูลรูปแบบที่ V เมื่อ K = 3, ρ = 0.2 และ n = 100 ด้วยสถิติทดสอบ t

ค่าที่วัด	t	df	Sig. (2-tailed)	ผลต่างค่าเฉลี่ย	ผลต่าง Std. Error
ค่า Purity	18.1625	1998.0000	**0.0000**	0.0166	0.0009
ค่า Rand statistic	17.7548	1985.6157	**0.0000**	0.0180	0.0010
ค่า Jaccard coefficient	18.1381	1974.5300	**0.0000**	0.0374	0.0021
ค่า Average silhouette width	152.9839	1916.3460	**0.0000**	0.1540	0.0010

ตารางที่ 4.56 ผลทดสอบความแตกต่างระหว่างค่าเฉลี่ยของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR และระยะห่างแบบ KR&N สำหรับข้อมูลรูปแบบที่ V เมื่อ K = 3, ρ = 0.8 และ n = 20 ด้วยสถิติทดสอบ t

ค่าที่วัด	t	df	Sig. (2-tailed)	ผลต่างค่าเฉลี่ย	ผลต่าง Std. Error
ค่า Purity	7.1283	1998.0000	**0.0000**	0.0170	0.0024
ค่า Rand statistic	13.5189	1830.9777	**0.0000**	0.0363	0.0027
ค่า Jaccard coefficient	11.3444	1987.0235	**0.0000**	0.0444	0.0039
ค่า Average silhouette width	46.5844	1968.1923	**0.0000**	0.1338	0.0029

ตารางที่ 4.57 ผลทดสอบความแตกต่างระหว่างค่าเฉลี่ยของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR และระยะห่างแบบ KR&N สำหรับข้อมูลรูปแบบที่ V เมื่อ K = 3, ρ = 0.8 และ n = 100 ด้วยสถิติทดสอบ t

ค่าที่วัด	t	df	Sig. (2-tailed)	ผลต่างค่าเฉลี่ย	ผลต่าง Std. Error
ค่า Purity	8.0355	1975.9082	**0.0000**	0.0088	0.0011
ค่า Rand statistic	12.1835	1428.1949	**0.0000**	0.0203	0.0017
ค่า Jaccard coefficient	12.8857	1769.6266	**0.0000**	0.0257	0.0020
ค่า Average silhouette width	129.3022	1963.5919	**0.0000**	0.1843	0.0014

4.1.5.2 ผลทดสอบความแตกต่างการวิเคราะห์กลุ่มข้อมูลรูปแบบที่ V เมื่อ K = 5

เมื่อ K = 5 ทดสอบความแตกต่างระหว่างค่าเฉลี่ยของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างระหว่างระยะห่างของ KR และระยะห่างแบบ KR&N ของข้อมูลต่อไปนี้

ข้อมูลรูปแบบที่ V เมื่อ K = 5, ρ = 0.2 และ n = 20

ข้อมูลรูปแบบที่ V เมื่อ K = 5, ρ = 0.2 และ n = 100

พบว่า ค่า Purity ค่า Rand statistic ค่า Jaccard coefficient และค่า Average silhouette width จากการวิเคราะห์กลุ่ม มี Sig. (2-tailed) น้อยกว่า 0.05 ดังตารางที่ 4.58 และ 4.59 จึงปฏิเสธสมมติฐานหลัก ดังนั้นการวิเคราะห์กลุ่มด้วยระยะห่างของ KR มีประสิทธิภาพและมีค่า Average silhouette width แตกต่างจากการวิเคราะห์กลุ่มด้วยระยะห่างแบบ KR&N ที่ระดับนัยสำคัญ 0.05

ตารางที่ 4.58 ผลทดสอบความแตกต่างระหว่างค่าเฉลี่ยของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR และระยะห่างแบบ KR&N สำหรับข้อมูลรูปแบบที่ V เมื่อ K = 5, ρ = 0.2 และ n = 20 ด้วยสถิติทดสอบ t

ค่าที่วัด	t	df	Sig. (2-tailed)	ผลต่างค่าเฉลี่ย	ผลต่าง Std. Error
ค่า Purity	9.5317	1998.0000	**0.0000**	0.0166	0.0017
ค่า Rand statistic	8.4005	1998.0000	**0.0000**	0.0091	0.0011
ค่า Jaccard coefficient	9.3031	1990.7859	**0.0000**	0.0288	0.0031
ค่า Average silhouette width	-2.2297	1998.0000	**0.0259**	-0.0043	0.0019

ตารางที่ 4.59 ผลทดสอบความแตกต่างระหว่างค่าเฉลี่ยของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR และระยะห่างแบบ KR&N สำหรับข้อมูลรูปแบบที่ V เมื่อ K = 5, ρ = 0.2 และ n = 100 ด้วยสถิติทดสอบ t

ค่าที่วัด	t	df	Sig. (2-tailed)	ผลต่างค่าเฉลี่ย	ผลต่าง Std. Error
ค่า Purity	6.8918	1985.2295	**0.0000**	0.0059	0.0009
ค่า Rand statistic	3.2505	1966.9205	**0.0012**	0.0017	0.0005
ค่า Jaccard coefficient	5.8130	1975.8359	**0.0000**	0.0085	0.0015
ค่า Average silhouette width	-2.8392	1998.0000	**0.0046**	-0.0025	0.0009

ข้อมูลรูปแบบที่ V เมื่อ $K = 5$, $\rho = 0.8$ และ $n = 20$

การทดสอบความแตกต่างของค่าเฉลี่ยของค่าต่าง ๆ ด้วยค่าสถิติทดสอบ t พบว่า ค่า Purity ค่า Rand statistic ค่า Jaccard coefficient และค่า Average silhouette width จากการวิเคราะห์กลุ่ม มี Sig. (2-tailed) มากกว่า 0.05 ดังตารางที่ 4.60 จึงไม่สามารถปฏิเสธสมมติฐานหลัก ดังนั้นการวิเคราะห์กลุ่มด้วยระยะห่างของ KR มีประสิทธิภาพและมีค่า Average silhouette width ไม่แตกต่างจากการวิเคราะห์กลุ่มด้วยระยะห่างแบบ KR&N

ตารางที่ 4.60 ผลทดสอบความแตกต่างระหว่างค่าเฉลี่ยของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR และระยะห่างแบบ KR&N สำหรับข้อมูลรูปแบบที่ V เมื่อ $K = 5$, $\rho = 0.8$ และ $n = 20$ ด้วยสถิติทดสอบ t

ค่าที่วัด	t	df	Sig. (2-tailed)	ผลต่างค่าเฉลี่ย	ผลต่าง Std. Error
ค่า Purity	1.3371	1998.0000	0.1814	0.0026	0.0020
ค่า Rand statistic	1.5439	1998.0000	0.1228	0.0016	0.0011
ค่า Jaccard coefficient	1.6729	1998.0000	0.0945	0.0043	0.0026
ค่า Average silhouette width	-0.2395	1998.0000	0.8107	-0.0005	0.0021

ข้อมูลรูปแบบที่ V เมื่อ $K = 5$, $\rho = 0.8$ และ $n = 100$

การทดสอบความแตกต่างของค่าเฉลี่ยของค่าต่าง ๆ ด้วยค่าสถิติทดสอบ t ดังตารางที่ 4.61 พบว่า ค่า Purity ค่า Jaccard coefficient และค่า Average silhouette width จากการวิเคราะห์กลุ่ม มี Sig. (2-tailed) มากกว่า 0.05 จึงไม่สามารถปฏิเสธสมมติฐานหลัก แต่ค่า Rand statistic จากการวิเคราะห์กลุ่ม มี Sig. (2-tailed) น้อยกว่า 0.05 จึงปฏิเสธสมมติฐานหลัก เนื่องจากเห็นว่าค่า Rand statistic แตกต่างจากค่าอื่น ๆ เพียงค่าเดียวเท่านั้น ดังนั้นจึงสรุปว่าการวิเคราะห์กลุ่มด้วยระยะห่างของ KR มีประสิทธิภาพและมีค่า Average silhouette width ไม่แตกต่างจากการวิเคราะห์กลุ่มด้วยระยะห่างแบบ KR&N ในภาพรวม

ตารางที่ 4.61 ผลทดสอบความแตกต่างระหว่างค่าเฉลี่ยของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR และระยะห่างแบบ KR&N สำหรับข้อมูลรูปแบบที่ V เมื่อ K = 5, ρ = 0.8 และ n = 100 ด้วยสถิติทดสอบ t

ค่าที่วัด	t	df	Sig. (2-tailed)	ผลต่างค่าเฉลี่ย	ผลต่าง Std. Error
ค่า Purity	-1.6225	1998.0000	0.1049	-0.0014	0.0008
ค่า Rand statistic	-3.0809	1998.0000	**0.0021**	-0.0014	0.0005
ค่า Jaccard coefficient	-1.8020	1998.0000	0.0717	-0.0020	0.0011
ค่า Average silhouette width	-0.3836	1998.0000	0.7013	-0.0004	0.0009

4.1.6 ผลการวิเคราะห์กลุ่มข้อมูลรูปแบบที่ VI

ผู้วิจัยทำการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ 2 วิธี ได้แก่ ระยะห่างของ KR และระยะห่างของ P สำหรับข้อมูลรูปแบบที่ VI เนื่องจากเป็นข้อมูลที่ประกอบไปด้วยตัวแปรอันดับเพียงประเภทเดียว จำนวน 3 ตัวแปร และคำนวณค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของ ค่า Purity ค่า Rand statistic ค่า Jaccard coefficient และค่า Average silhouette width ได้ผลดังตารางที่ 4.6

ผลการเปรียบเทียบประสิทธิภาพโดยเฉลี่ยในการวิเคราะห์กลุ่มด้วยมาตรวัดระยะห่างของ KR และระยะห่างของ P พบว่า เมื่อ n= 20 การวิเคราะห์กลุ่มด้วยระยะห่างของ P ให้ค่าเฉลี่ยของ ค่า Purity ค่า Rand statistic และค่า Jaccard coefficient มากกว่าการวิเคราะห์กลุ่มด้วยระยะห่างของ KR ทุกกรณี (ยกเว้นกรณีที่ K = 5 และ ρ = 0.8) แต่เมื่อ n= 100 การวิเคราะห์กลุ่มด้วยระยะห่างทั้ง 2 วิธีนี้ ให้ค่าเฉลี่ยของ ค่า Purity ค่า Rand statistic และค่า Jaccard coefficient เท่ากันทุกกรณี ดังตารางที่ 4.62

นอกจากนี้ ค่าเฉลี่ยของค่า Average silhouette width จากการวิเคราะห์กลุ่มด้วยระยะห่างของ P มีค่ามากกว่าหรือเท่ากับ จากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR ในทุกกรณี ยกเว้นกรณีที่ K = 5 และ ρ = 0.8 ดังตารางที่ 4.62

ตารางที่ 4.62 ค่าเฉลี่ย (ส่วนเบี่ยงเบนมาตรฐาน) จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ VI

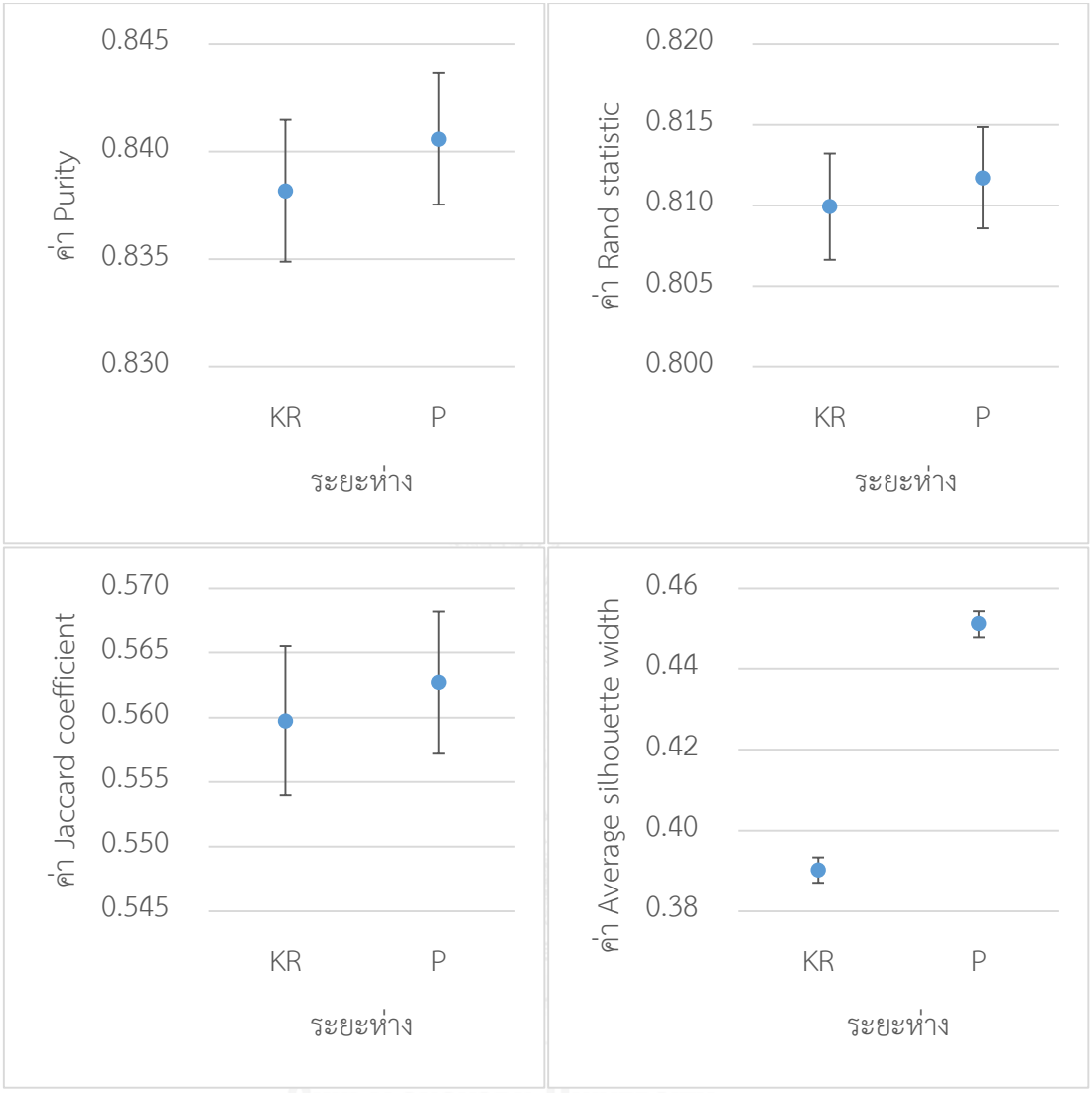
K	ρ	n	ค่าที่วัด	ระยะห่าง	
				KR	P
3	0.2	20	ค่า Purity	0.8382 (0.0532)	**0.8406** (0.0491)
			ค่า Rand statistic	0.8099 (0.0531)	**0.8117** (0.0505)
			ค่า Jaccard coefficient	0.5597 (0.0928)	**0.5627** (0.0888)
			ค่า Average silhouette width	0.3902 (0.0507)	**0.4511** (0.0538)
		100	ค่า Purity	0.8391 (0.0209)	0.8391 (0.0209)
			ค่า Rand statistic	0.8084 (0.0219)	0.8084 (0.0219)
			ค่า Jaccard coefficient	0.5599 (0.0373)	0.5599 (0.0373)
			ค่า Average silhouette width	0.3899 (0.0227)	**0.4514** (0.0243)
	0.8	20	ค่า Purity	0.7790 (0.0593)	**0.7821** (0.0568)
			ค่า Rand statistic	0.7467 (0.0611)	**0.7533** (0.0539)
			ค่า Jaccard coefficient	0.4593 (0.0883)	**0.4667** (0.0824)
			ค่า Average silhouette width	0.5603 (0.0561)	**0.6319** (0.0585)
		100	ค่า Purity	0.7825 (0.0231)	0.7825 (0.0231)
			ค่า Rand statistic	0.7521 (0.0220)	0.7521 (0.0220)
			ค่า Jaccard coefficient	0.4679 (0.0327)	0.4679 (0.0327)
			ค่า Average silhouette width	0.5576 (0.0253)	**0.6356** (0.0254)
5	0.2	20	ค่า Purity	0.7794 (0.0566)	**0.7825** (0.0537)
			ค่า Rand statistic	0.8488 (0.0334)	**0.8510** (0.0316)
			ค่า Jaccard coefficient	0.4501 (0.0818	**0.4545** (0.0791)
			ค่า Average silhouette width	0.3596 (0.0412)	**0.3607** (0.0411)
		100	ค่า Purity	0.7953 (0.0179)	0.7953 (0.0179)
			ค่า Rand statistic	0.8571 (0.0107)	0.8571 (0.0107)
			ค่า Jaccard coefficient	0.4740 (0.0288)	0.4740 (0.0288)
			ค่า Average silhouette width	0.3570 (0.0175)	0.3570 (0.0176)

ตารางที่ 4.62 (ต่อ) ค่าเฉลี่ย (ส่วนเบี่ยงเบนมาตรฐาน) จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ VI

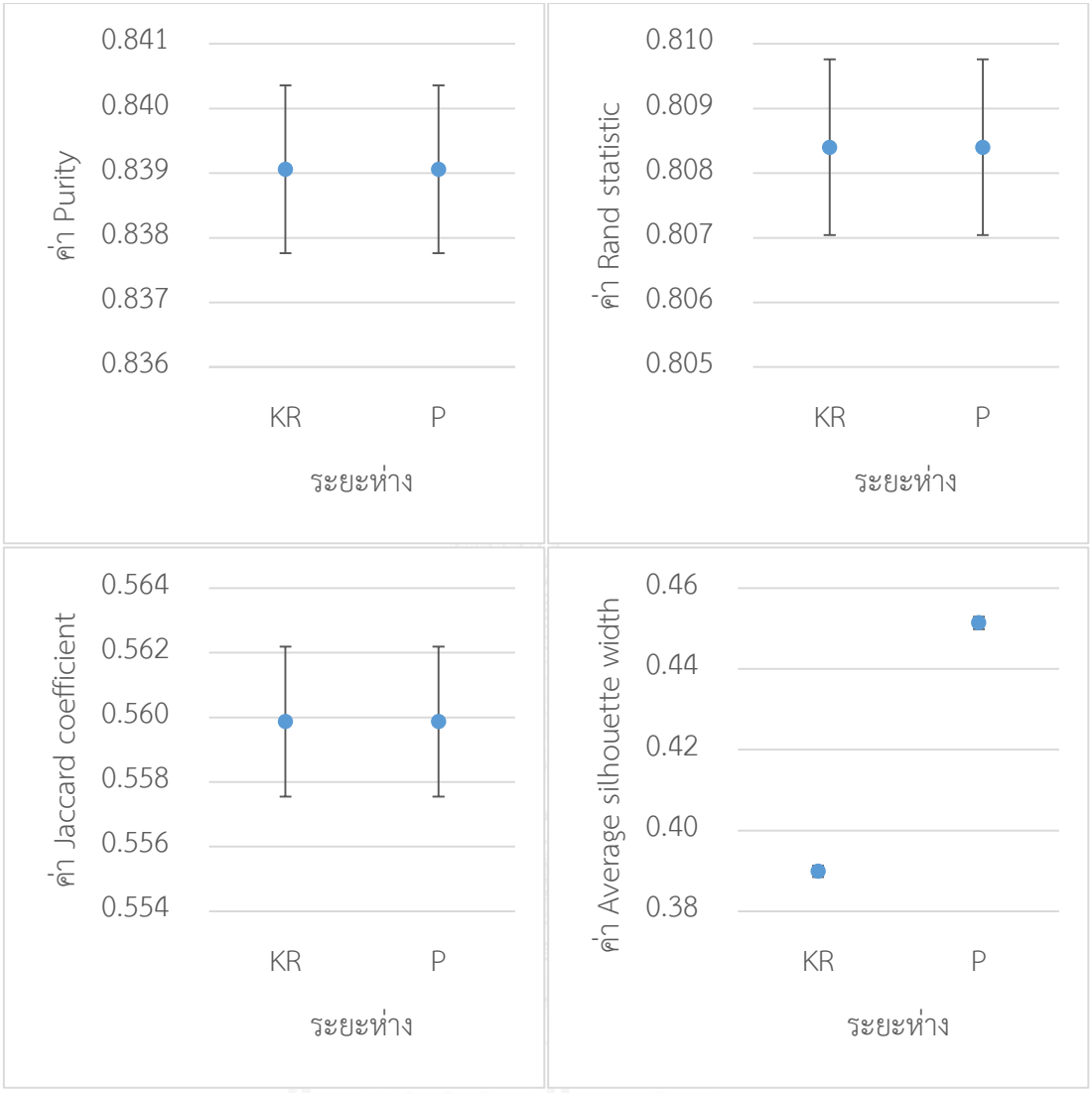
K	ρ	n	ค่าที่วัด	ระยะห่าง	
				KR	P
5	0.8	20	ค่า Purity	0.7184 (0.0448)	0.7184 (0.0448)
			ค่า Rand statistic	0.8180 (0.0230)	0.8180 (0.0230)
			ค่า Jaccard coefficient	0.3642 (0.0552)	0.3642 (0.0552)
			ค่า Average silhouette width	**0.5832** (0.0455)	0.5819 (0.0462)
		100	ค่า Purity	0.7184 (0.0205)	0.7184 (0.0205)
			ค่า Rand statistic	0.8158 (0.0105)	0.8158 (0.0105)
			ค่า Jaccard coefficient	0.3676 (0.0247)	0.3676 (0.0247)
			ค่า Average silhouette width	**0.5870** (0.0212)	0.5867 (0.0216)

เมื่อพิจารณากราฟช่วงความเชื่อมั่น 95% ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR และระยะห่างของ P สำหรับข้อมูลรูปแบบที่ VI พบว่าค่าเฉลี่ยของค่า Purity ค่า Rand statistic และค่า Jaccard coefficient จากการวิเคราะห์กลุ่มด้วยระยะห่างทั้ง 2 วิธีนี้ มีค่าใกล้เคียงกัน โดยมีส่วนที่ช่วงความเชื่อมั่น 95% ซ้อนทับกันในทุกกรณี แสดงว่าการวิเคราะห์กลุ่มด้วยระยะห่างทั้ง 2 วิธีนี้มีประสิทธิภาพโดยเฉลี่ยเท่ากันหรือใกล้เคียงกันในทุกกรณี ดังรูปที่ 4.41 ถึง 4.48

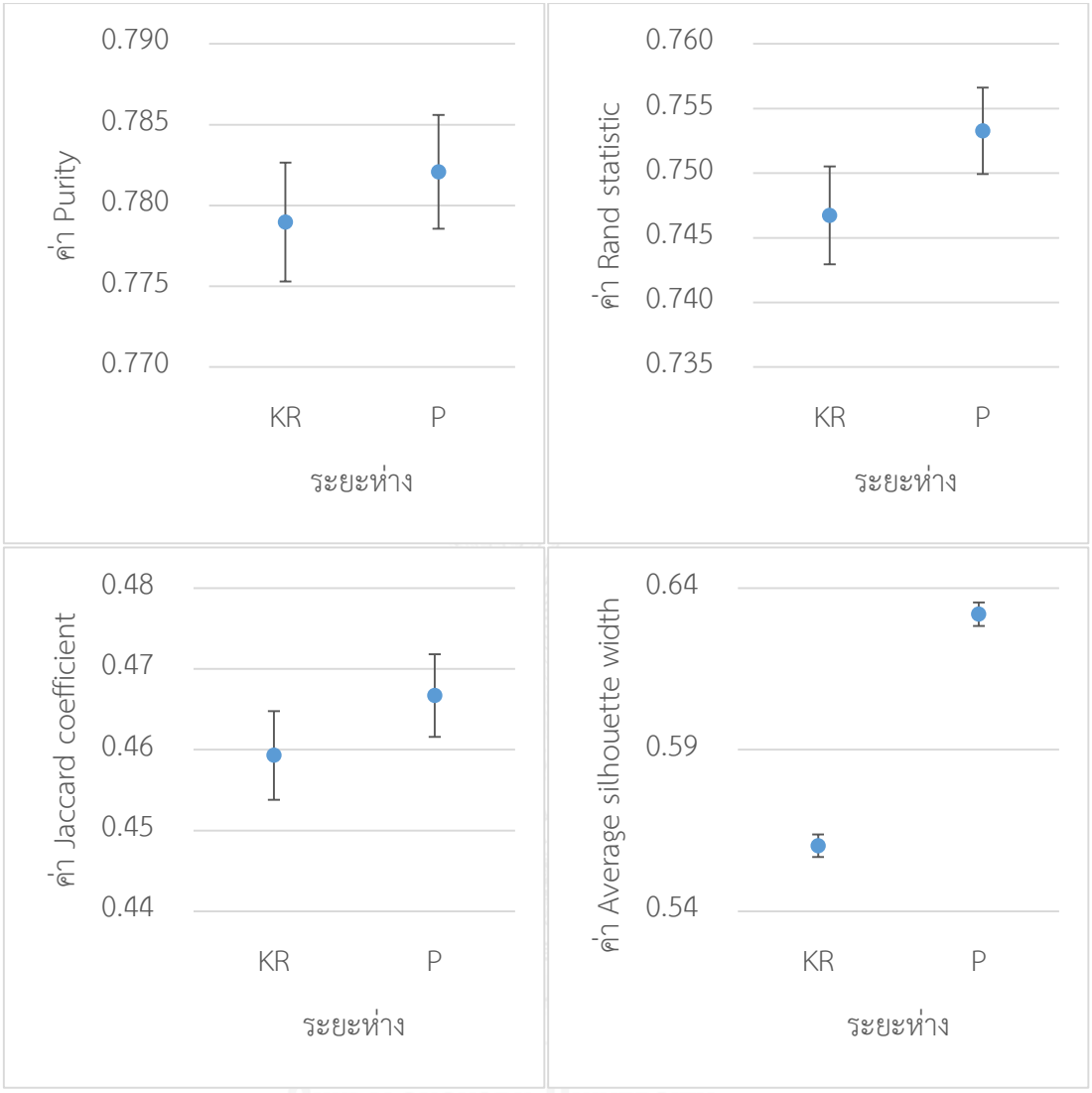
ขณะที่กราฟช่วงความเชื่อมั่น 95% ของค่า Average silhouette width พบว่าเมื่อ K = 3 ค่าเฉลี่ยของค่า Average silhouette width จากการวิเคราะห์กลุ่มด้วยระยะห่างของ P มีค่ามากกว่าจากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR โดยไม่มีส่วนที่ช่วงความเชื่อมั่น 95% ซ้อนทับกัน ดังรูปที่ 4.41 ถึง 4.44 แต่เมื่อ K = 5 ค่าเฉลี่ยของค่า Average silhouette width จากการวิเคราะห์กลุ่มด้วยระยะห่างของ P และระยะห่างของ KR มีค่าใกล้เคียงกันและมีส่วนที่ช่วงความเชื่อมั่น 95% ซ้อนทับกัน ดังรูปที่ 4.45 และ 4.48



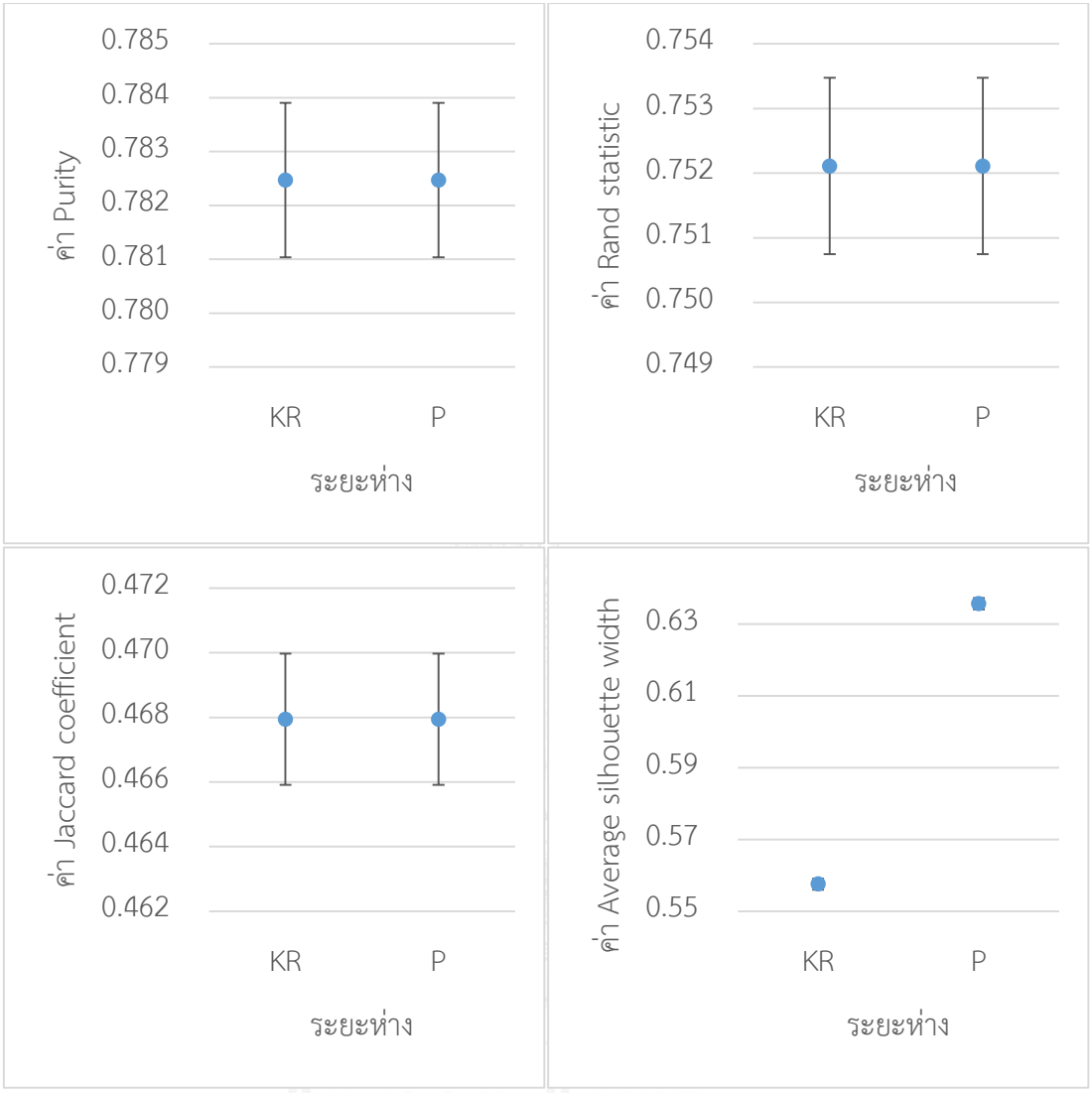
รูปที่ 4.41 กราฟช่วงความเชื่อมั่น 95% ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ VI เมื่อ $K = 3$, $\rho = 0.2$ และ $n = 20$



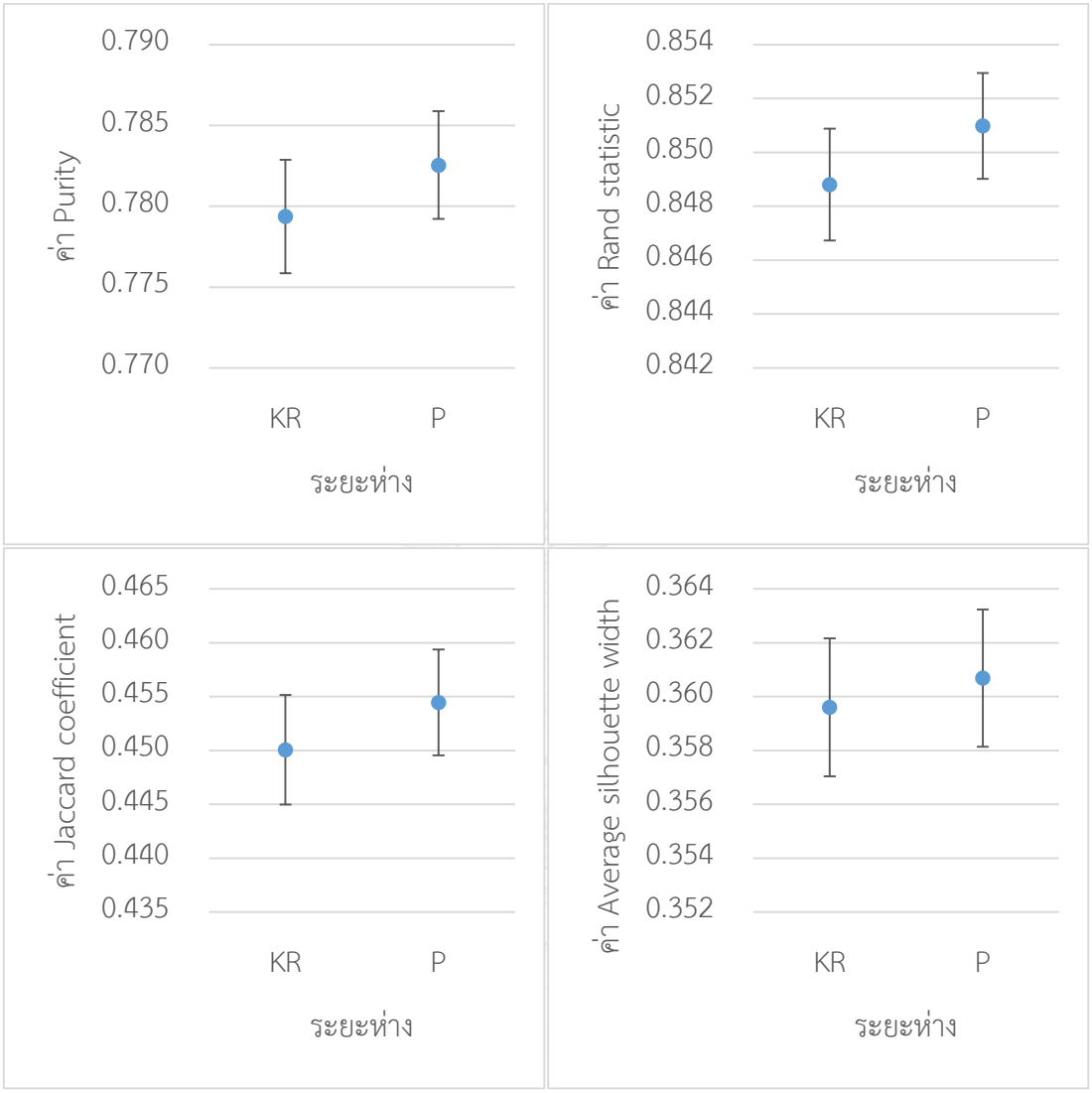
รูปที่ 4.42 กราฟช่วงความเชื่อมั่น 95% ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ VI เมื่อ $K = 3$, $\rho = 0.2$ และ $n = 100$



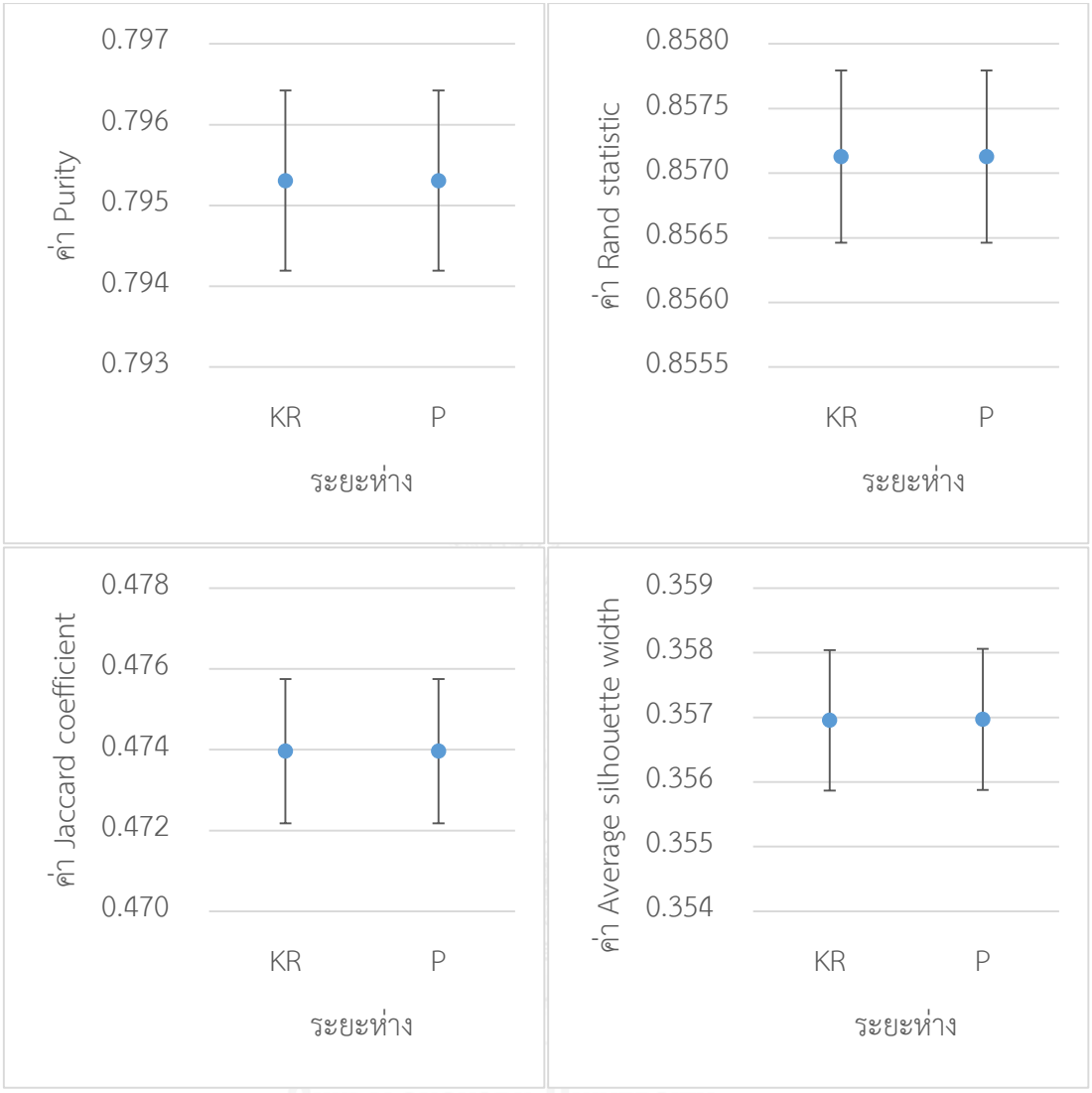
รูปที่ 4.43 กราฟช่วงความเชื่อมั่น 95% ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ VI เมื่อ $K = 3$, $\rho = 0.8$ และ $n = 20$



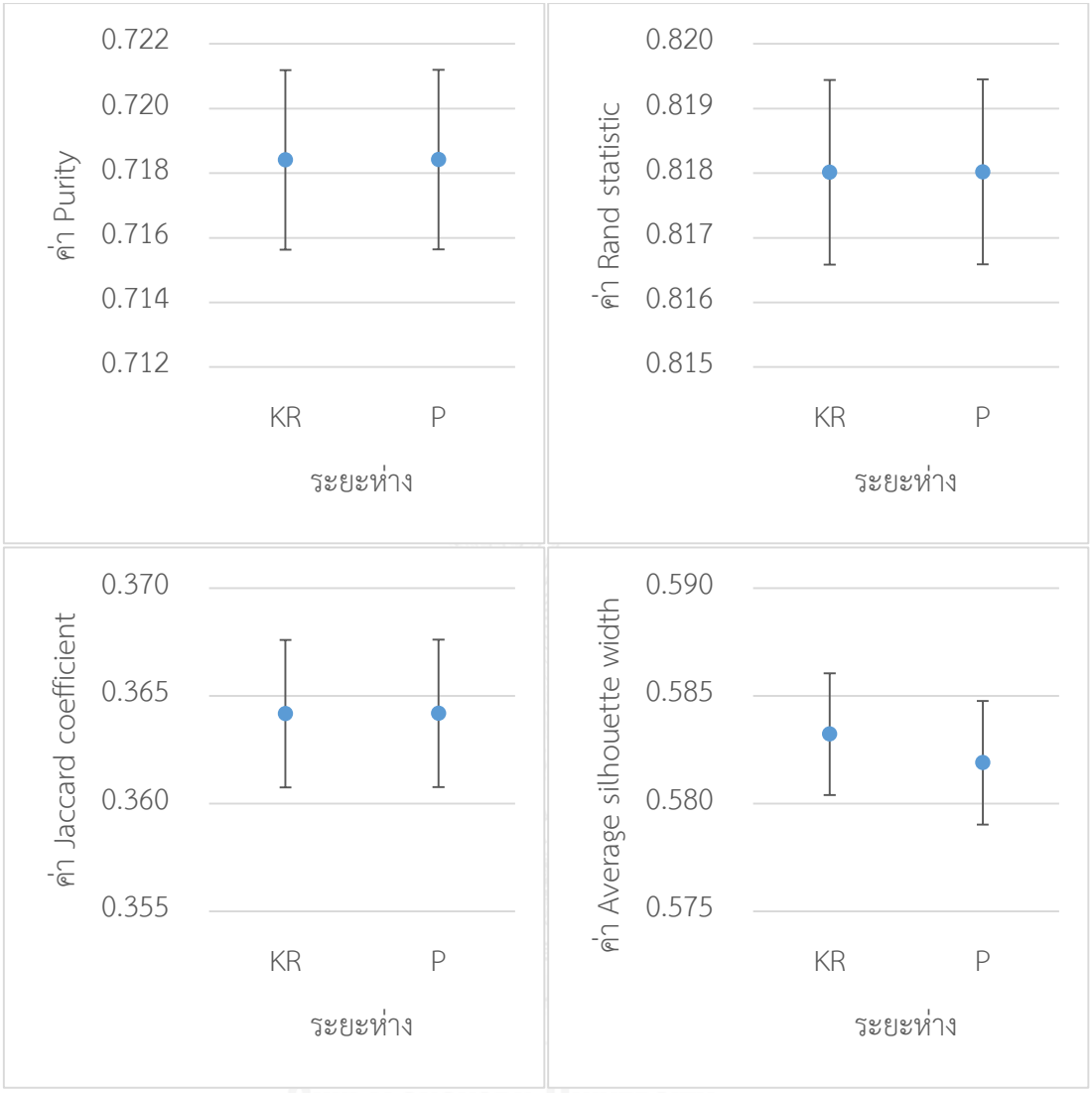
รูปที่ 4.44 กราฟช่วงความเชื่อมั่น 95% ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ VI เมื่อ $K = 3$, $\rho = 0.8$ และ $n = 100$



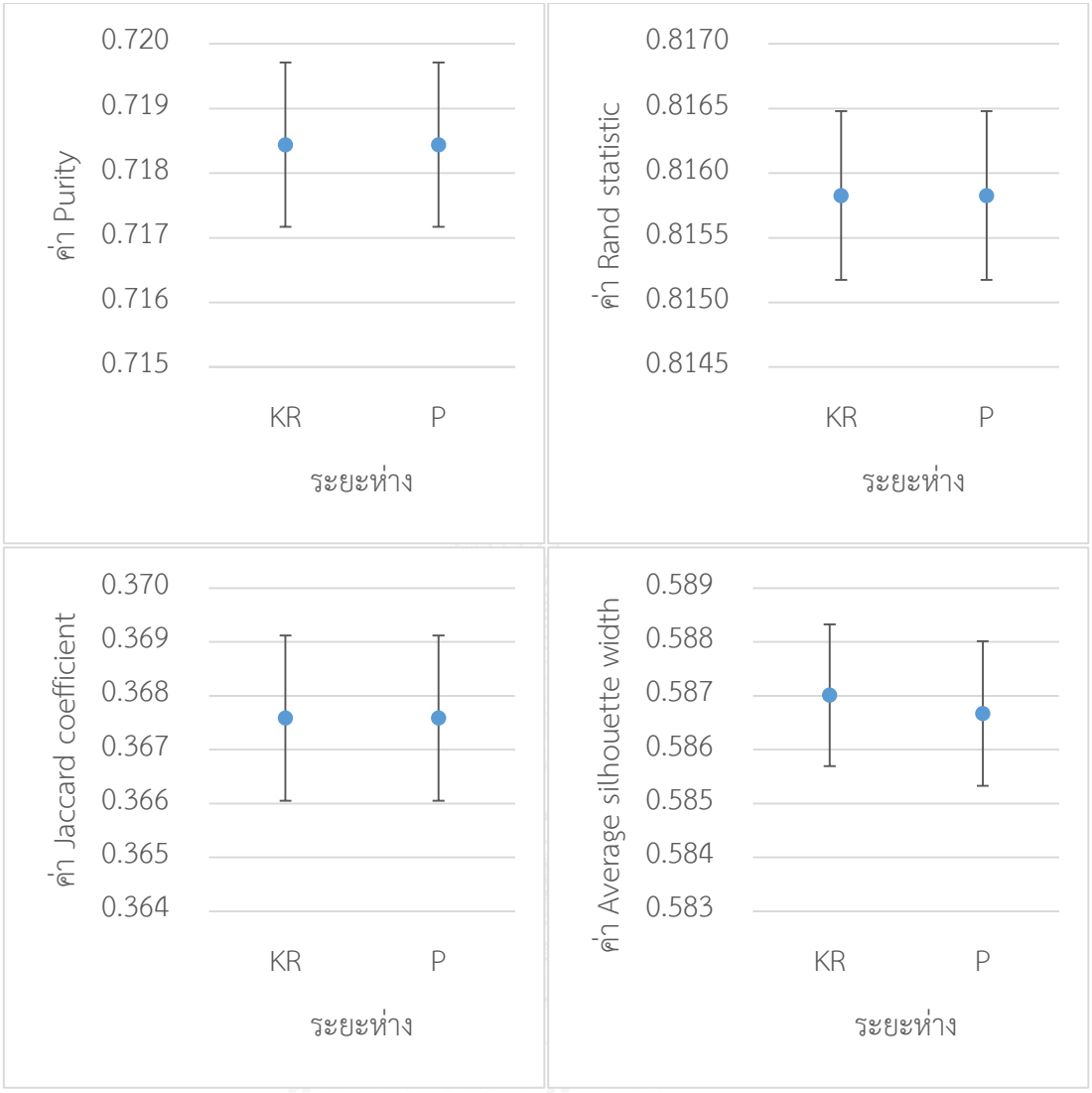
รูปที่ 4.45 กราฟช่วงความเชื่อมั่น 95% ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ VI เมื่อ $K = 5$, $\rho = 0.2$ และ $n = 20$



รูปที่ 4.46 กราฟช่วงความเชื่อมั่น 95% ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ VI เมื่อ $K = 5$, $\rho = 0.2$ และ $n = 100$



รูปที่ 4.47 กราฟช่วงความเชื่อมั่น 95% ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ VI เมื่อ $K = 5$, $\rho = 0.8$ และ $n = 20$



รูปที่ 4.48 กราฟช่วงความเชื่อมั่น 95% ของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูลรูปแบบที่ VI เมื่อ $K = 5$, $\rho = 0.8$ และ $n = 100$

เนื่องจากการพิจารณากราฟช่วงความเชื่อมั่น 95% พบว่าค่าเฉลี่ยของค่า Purity ค่า Rand statistic ค่า Jaccard coefficient และค่า Average silhouette width มีค่าใกล้เคียงกันในบางกรณี ทั้งยังมีส่วนของช่วงความเชื่อมั่น 95% ทับซ้อนกัน จึงไม่สามารถสรุปได้ว่าการวิเคราะห์กลุ่มด้วยระยะห่างแบบใดมีประสิทธิภาพดีที่สุด ผู้วิจัยจึงทดสอบความแตกต่างระหว่างค่าเฉลี่ยของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างแต่ละคู่ ด้วยสถิติทดสอบ t และแบ่งการพิจารณาข้อมูลรูปแบบที่ V เป็นกรณีใหญ่ 2 กรณีตามจำนวนกลุ่มข้อมูล คือ กรณีที่ $K = 3$ และ $K = 5$

4.1.6.1 ผลทดสอบความแตกต่างการวิเคราะห์กลุ่มข้อมูลรูปแบบที่ VI เมื่อ K = 3

ศึกษาประสิทธิภาพการวิเคราะห์กลุ่ม โดยทดสอบความแตกต่างของค่าเฉลี่ยของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างระหว่างระยะห่างของ KR และระยะห่างของ P ด้วยค่าสถิติทดสอบ t สำหรับข้อมูลที่แตกต่างกันดังต่อไปนี้ $K = 3$

ข้อมูลรูปแบบที่ VI เมื่อ K = 3, ρ = 0.2 และ n = 20

ข้อมูลรูปแบบที่ VI เมื่อ K = 3, ρ = 0.2 และ n = 100

ข้อมูลรูปแบบที่ VI เมื่อ K = 3, ρ = 0.8 และ n = 20

ข้อมูลรูปแบบที่ VI เมื่อ K = 3, ρ = 0.8 และ n = 100

ได้ผลดังตารางที่ 4.63 ถึง 4.66 พบว่า ค่า Purity ค่า Rand statistic และค่า Jaccard coefficient จากการวิเคราะห์กลุ่มข้อมูลในทุกกรณี มี Sig. (2-talied) มากกว่า 0.05 จึงไม่สามารถปฏิเสธสมมติฐานหลักในทุกกรณี ยกเว้นค่า Rand statistic กรณี K = 3, ρ = 0.8 และ n = 20 พบว่า มี .Sig (2-talied) เท่ากับ 0.0112 ซึ่งน้อยกว่า 0.05 จึงปฏิเสธสมมติฐานหลัก แต่เนื่องจากค่า Purity และค่า Jaccard coefficient มีผลไปทางเดียวกันและสนับสนุนกัน ซึ่งขัดแย้งกับค่า Rand statistic เพียงค่าเดียวเท่านั้น ดังนั้นโดยภาพรวมจึงสรุปได้ว่า การวิเคราะห์กลุ่มด้วยระยะห่างของ KR มีประสิทธิภาพไม่แตกต่างจากการวิเคราะห์กลุ่มด้วยระยะห่างของ P

นอกจากนี้ ค่า Average silhouette width มี Sig. (2-talied) น้อยกว่า 0.05 จึงปฏิเสธสมมติฐานหลัก ดังนั้นการวิเคราะห์กลุ่มด้วยระยะห่างของ KR ให้ค่า Average silhouette width แตกต่างจากการวิเคราะห์กลุ่มด้วยระยะห่างของ P

ตารางที่ 4.63 ผลทดสอบความแตกต่างระหว่างค่าเฉลี่ยของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR และระยะห่างของ P สำหรับข้อมูลรูปแบบที่ VI เมื่อ K = 3, ρ = 0.2 และ n = 20 ด้วยสถิติทดสอบ t

ค่าที่วัด	t	df	Sig. (2-tailed)	ผลต่างค่าเฉลี่ย	ผลต่าง Std. Error
ค่า Purity	-1.0485	1998.0000	0.2945	-0.0024	0.0023
ค่า Rand statistic	-0.7692	1998.0000	0.4419	-0.0018	0.0023
ค่า Jaccard coefficient	-0.7320	1998.0000	0.4642	-0.0030	0.0041
ค่า Average silhouette width	-26.0635	1998.0000	**0.0000**	-0.0609	0.0023

ตารางที่ 4.64 ผลทดสอบความแตกต่างระหว่างค่าเฉลี่ยของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR และระยะห่างของ P สำหรับข้อมูลรูปแบบที่ VI เมื่อ K = 3, ρ = 0.2 และ n = 100 ด้วยสถิติทดสอบ t

ค่าที่วัด	t	df	Sig. (2-tailed)	ผลต่างค่าเฉลี่ย	ผลต่าง Std. Error
ค่า Purity	0.0000	1998.0000	1.0000	0.0000	0.0009
ค่า Rand statistic	0.0000	1998.0000	1.0000	0.0000	0.0010
ค่า Jaccard coefficient	0.0000	1998.0000	1.0000	0.0000	0.0017
ค่า Average silhouette width	-58.5013	1998.0000	**0.0000**	-0.0615	0.0011

ตารางที่ 4.65 ผลทดสอบความแตกต่างระหว่างค่าเฉลี่ยของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR และระยะห่างของ P สำหรับข้อมูลรูปแบบที่ VI เมื่อ K = 3, ρ = 0.8 และ n = 20 ด้วยสถิติทดสอบ t

ค่าที่วัด	t	df	Sig. (2-tailed)	ผลต่างค่าเฉลี่ย	ผลต่าง Std. Error
ค่า Purity	-1.1942	1998.0000	0.2325	-0.0031	0.0026
ค่า Rand statistic	-2.5378	1966.9981	**0.0112**	-0.0065	0.0026
ค่า Jaccard coefficient	-1.9372	1988.6867	0.0529	-0.0074	0.0038
ค่า Average silhouette width	-27.9685	1998.0000	**0.0000**	-0.0717	0.0026

ตารางที่ 4.66 ผลทดสอบความแตกต่างระหว่างค่าเฉลี่ยของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR และระยะห่างของ P สำหรับข้อมูลรูปแบบที่ VI เมื่อ K = 3, ρ = 0.8 และ n = 100 ด้วยสถิติทดสอบ t

ค่าที่วัด	t	df	Sig. (2-tailed)	ผลต่างค่าเฉลี่ย	ผลต่าง Std. Error
ค่า Purity	0.0000	1998.0000	1.0000	0.0000	0.0010
ค่า Rand statistic	0.0000	1998.0000	1.0000	0.0000	0.0010
ค่า Jaccard coefficient	0.0000	1998.0000	1.0000	0.0000	0.0015
ค่า Average silhouette width	-68.8460	1998.0000	**0.0000**	-0.0781	0.0011

4.1.6.2 ผลทดสอบความแตกต่างการวิเคราะห์กลุ่มข้อมูลรูปแบบที่ VI เมื่อ K = 5

ศึกษาประสิทธิภาพการวิเคราะห์กลุ่มข้อมูลต่าง ๆ โดยทดสอบความแตกต่างของค่าเฉลี่ยของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างระหว่างระยะห่างของ KR และระยะห่างของ P ด้วยค่าสถิติทดสอบ t สำหรับข้อมูลดังกรณีต่อไปนี้

ข้อมูลรูปแบบที่ VI เมื่อ K = 5, ρ = 0.2 และ n = 20

ข้อมูลรูปแบบที่ VI เมื่อ K = 5, ρ = 0.2 และ n = 100

ข้อมูลรูปแบบที่ VI เมื่อ K = 5, ρ = 0.8 และ n = 20

ข้อมูลรูปแบบที่ VI เมื่อ K = 5, ρ = 0.8 และ n = 100

ได้ผลดังตารางที่ 4.67 ถึง 4.70 พบว่า ค่า Purity ค่า Rand statistic ค่า Jaccard coefficient และค่า Average silhouette width จากการวิเคราะห์กลุ่มข้อมูลในทุกกรณี มี Sig. (2-talied) มากกว่า 0.05 จึงไม่สามารถปฏิเสธสมมติฐานหลัก ดังนั้น การวิเคราะห์กลุ่มด้วยระยะห่างของ KR มีประสิทธิภาพและมีค่า Average silhouette width ไม่แตกต่างจากการวิเคราะห์กลุ่มด้วยระยะห่างของ P

ตารางที่ 4.67 ผลทดสอบความแตกต่างระหว่างค่าเฉลี่ยของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR และระยะห่างของ P สำหรับข้อมูลรูปแบบที่ VI เมื่อ K = 5, ρ = 0.2 และ n = 20 ด้วยสถิติทดสอบ t

ค่าที่วัด	t	df	Sig. (2-tailed)	ผลต่างค่าเฉลี่ย	ผลต่าง Std. Error
ค่า Purity	-1.2890	1998.0000	0.1980	-0.0032	0.0025
ค่า Rand statistic	-1.4980	1998.0000	0.1340	-0.0022	0.0015
ค่า Jaccard coefficient	-1.2180	1998.0000	0.2230	-0.0044	0.0036
ค่า Average silhouette width	-0.5920	1998.0000	0.5540	-0.0011	0.0018

ตารางที่ 4.68 ผลทดสอบความแตกต่างระหว่างค่าเฉลี่ยของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR และระยะห่างของ P สำหรับข้อมูลรูปแบบที่ VI เมื่อ K = 5, ρ = 0.2 และ n = 100 ด้วยสถิติทดสอบ t

ค่าที่วัด	t	df	Sig. (2-tailed)	ผลต่างค่าเฉลี่ย	ผลต่าง Std. Error
ค่า Purity	0.0000	1998.0000	1.0000	0.0000	0.0008
ค่า Rand statistic	0.0000	1998.0000	1.0000	0.0000	0.0005
ค่า Jaccard coefficient	0.0000	1998.0000	1.0000	0.0000	0.0013
ค่า Average silhouette width	-0.0200	1998.0000	0.9841	0.0000	0.0008

ตารางที่ 4.69 ผลทดสอบความแตกต่างระหว่างค่าเฉลี่ยของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR และระยะห่างของ P สำหรับข้อมูลรูปแบบที่ VI เมื่อ K = 5, ρ = 0.8 และ n = 20 ด้วยสถิติทดสอบ t

ค่าที่วัด	t	df	Sig. (2-tailed)	ผลต่างค่าเฉลี่ย	ผลต่าง Std. Error
ค่า Purity	-0.0050	1998.0000	0.9960	0.0000	0.0020
ค่า Rand statistic	-0.0063	1998.0000	0.9950	0.0000	0.0010
ค่า Jaccard coefficient	-0.0056	1998.0000	0.9955	0.0000	0.0025
ค่า Average silhouette width	0.6479	1998.0000	0.5171	0.0013	0.0021

ตารางที่ 4.70 ผลทดสอบความแตกต่างระหว่างค่าเฉลี่ยของค่าต่าง ๆ จากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR และระยะห่างของ P สำหรับข้อมูลรูปแบบที่ VI เมื่อ K = 5, ρ = 0.8 และ n = 100 ด้วยสถิติทดสอบ t

ค่าที่วัด	t	df	Sig. (2-tailed)	ผลต่างค่าเฉลี่ย	ผลต่าง Std. Error
ค่า Purity	0.0000	1998.0000	1.0000	0.0000	0.0009
ค่า Rand statistic	0.0000	1998.0000	1.0000	0.0000	0.0005
ค่า Jaccard coefficient	0.0000	1998.0000	1.0000	0.0000	0.0011
ค่า Average silhouette width	0.3545	1998.0000	0.7230	0.0003	0.0010

4.2 ผลการวิเคราะห์กลุ่มข้อมูลจริง

ผู้วิจัยศึกษาประสิทธิภาพระยะห่างแบบต่าง ๆ จากการวิเคราะห์กลุ่มข้อมูลจริง 1 ชุด คือ ข้อมูล Pittsburgh Bridges Data Set มีขนาดข้อมูลเท่ากับ 108 ข้อมูล ซึ่งเป็นข้อมูลแบบผสม โดยประกอบไปด้วยตัวแปรต่าง ๆ ดังนี้

1. ตัวแปรอิสระ		จำนวน	11	ตัวแปร
ได้แก่	ตัวแปรนามบัญญัติ	จำนวน	7	ตัวแปร
	- ตัวแปร RIVER	จำนวนประเภท	4	ประเภท
	- ตัวแปร LOCATION	จำนวนประเภท	54	ประเภท
	- ตัวแปร PURPOSE	จำนวนประเภท	4	ประเภท
	- ตัวแปร CLEAR-G	จำนวนประเภท	2	ประเภท
	- ตัวแปร T-OR-D	จำนวนประเภท	2	ประเภท
	- ตัวแปร MATERIAL	จำนวนประเภท	3	ประเภท
	- ตัวแปร REL-L	จำนวนประเภท	3	ประเภท
	ตัวแปรอันดับ	จำนวน	1	ตัวแปร
	- ตัวแปร SPAN	จำนวนอันดับ	3	อันดับ
	ตัวแปรเชิงปริมาณ	จำนวน	3	ตัวแปร
	- ตัวแปร ERECTED	มีค่าอยู่ระหว่าง	1818 ถึง 1986	
	- ตัวแปร LENGTH	มีค่าอยู่ระหว่าง	804 ถึง 4558	
	- ตัวแปร LANES	มีค่าอยู่ระหว่าง	1 ถึง 6	
2. ตัวแปรตาม		จำนวน	1	ตัวแปร
ได้แก่	ตัวแปรนามบัญญัติ	จำนวน	1	ตัวแปร
	- ตัวแปร TYPE	จำนวนประเภท	7	ประเภท

คำนวณค่า Purity ค่า Rand statistic ค่า Jaccard coefficient และค่า Average silhouette width ที่ได้จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ พบว่าการวิเคราะห์กลุ่มด้วยระยะห่างแบบ KR&N ให้ค่า Purity และค่า Jaccard coefficient สูงที่สุดเท่ากับ 0.5185 และ 0.2274 ตามลำดับ แต่ให้ค่า Rand statistic ต่ำที่สุดเท่ากับ 0.7459 โดยที่การวิเคราะห์กลุ่มด้วยระยะห่างแบบ P&N ให้ค่า Rand statistic สูงที่สุดเท่ากับ 0.7617 นอกจากนี้ค่า Average silhouette width จากการวิเคราะห์กลุ่มด้วยระยะห่างของ KR มีค่าสูงที่สุด และใกล้เคียงกับระยะห่างของ P ตามด้วยระยะห่างแบบ KR&N และระยะห่างแบบ P&N ตามลำดับ ดังตารางที่ 4.71

ตารางที่ 4.71 ผลการวัดประสิทธิภาพการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ สำหรับข้อมูล Pittsburgh Bridges Data Set

ค่าที่ได้	ระยะห่าง			
	KR	P	KR&N	P&N
ค่า Purity	0.4815	0.4630	**0.5185**	0.4907
ค่า Rand statistic	0.7541	0.7520	0.7459	**0.7617**
ค่า Jaccard coefficient	0.2075	0.2026	**0.2274**	0.2176
ค่า Average silhouette width	**0.2982**	0.2862	0.2586	0.2383

บทที่ 5
สรุปผลการวิจัยและข้อเสนอแนะ

การวิจัยครั้งนี้ มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพในการวิเคราะห์กลุ่มด้วยอัลกอริทึมจัดกลุ่มโดยรอบมีดอยด์ที่ใช้ระยะห่างแบบต่าง ๆ สำหรับข้อมูลแบบผสมที่ประกอบไปด้วยตัวแปรเชิงปริมาณ ตัวแปรนามบัญญัติ และตัวแปรอันดับ ซึ่งข้อมูลที่จำลองขึ้นมีลักษณะที่แตกต่างกันตามจำนวนกลุ่ม (K = 3, 5) ค่าสัมประสิทธิ์สหสัมพันธ์ (ρ = 0.2, 0.8) ขนาดกลุ่มข้อมูล (n = 20, 100) และชนิดและจำนวนตัวแปรอิสระ ได้แก่ ตัวแปรนามบัญญัติ ตัวแปรอันดับ และตัวแปรเชิงปริมาณ อย่างละ 3 ตัวแปร โดยตัวแปรนามบัญญัติและตัวแปรอันดับ มีจำนวนประเภทหรืออันดับเท่ากับ 5 ประเภทหรืออันดับ ซึ่งกำหนดให้ความน่าจะเป็นที่ประเภทหรืออันดับต่าง ๆ ของแต่ละตัวแปร เกิดขึ้นแตกต่างกันในแต่ละกลุ่ม และกำหนดค่าเฉลี่ยของตัวแปรเชิงปริมาณแตกต่างกันในแต่ละกลุ่มอีกด้วย เพื่อให้ข้อมูลมีความแตกต่างระหว่างกลุ่ม

ผู้วิจัยทำการวิเคราะห์กลุ่มข้อมูลที่จำลองขึ้น ซึ่งประกอบไปด้วยตัวแปรอิสระชนิดต่าง ๆ 6 รูปแบบ ดังตารางที่ 5.1

ตารางที่ 5.1 รูปแบบข้อมูลแบ่งตามประเภทตัวแปร

ประเภทตัวแปร	ข้อมูลรูปแบบที่					
	I	II	III	IV	V	VI
ตัวแปรนามบัญญัติ	✓	✓	✓	-	✓	-
ตัวแปรอันดับ	✓	✓	-	✓	-	✓
ตัวแปรเชิงปริมาณ	✓	-	✓	✓	-	-

และวิเคราะห์กลุ่มข้อมูลจริง 1 ชุด คือ ข้อมูล Pittsburgh Bridges Data Set ซึ่งเป็นข้อมูลแบบผสมประกอบไปด้วยตัวแปรทั้ง 3 ชนิด

มาตรวัดระยะห่างที่สนใจศึกษาเป็นระยะห่างสำหรับข้อมูลแบบผสม ซึ่งมี 4 วิธีดังนี้

1. ระยะห่างของ Kaufman and Rousseeuw (KR)
2. ระยะห่างของ Podani (P)
3. ระยะห่างแบบ Kaufman and Rousseeuw ร่วมกับ Noorbehbahani et al. (KR&N)
4. ระยะห่างแบบ Podani ร่วมกับ Noorbehbahani et al. (P&N)

ซึ่งระยะห่างแบบ KR&N และ P&N เป็นระยะห่างที่ผู้วิจัยนำเสนอขึ้นใหม่ โดยประยุกต์จาก ระยะห่างของ KR ระยะห่างของ P และระยะห่างสำหรับตัวแปรนามบัญญัติของ Noorbehbahani et al. เพื่อเพิ่มประสิทธิภาพในการวิเคราะห์กลุ่มสำหรับข้อมูลแบบผสมมากยิ่งขึ้น

5.1 สรุปผลการวิจัย

5.1.1 สรุปผลการวิเคราะห์กลุ่มข้อมูลที่จำลองขึ้น

ข้อมูลที่จำลองขึ้น มีทั้งกรณีที่เป็นข้อมูลแบบผสมที่ประกอบไปด้วยตัวแปรทั้ง 3 ชนิด และข้อมูลแบบผสมที่ประกอบไปด้วยตัวแปรบางชนิดเท่านั้น รวมทั้งสิ้น 6 รูปแบบ ในการสรุปผลการศึกษาจะพิจารณาประสิทธิภาพในการวิเคราะห์กลุ่มข้อมูลนี้ในภาพรวม ซึ่งมีความหมายครอบคลุมค่า Purity ค่า Rand statistic และค่า Jaccard coefficient เนื่องจากทั้ง 3 ค่านี้บ่งชี้ประสิทธิภาพการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ ไปในทิศทางเดียวกัน และแยกพิจารณาค่า Average silhouette width เนื่องจากไม่สามารถชี้วัดประสิทธิภาพการวิเคราะห์กลุ่มข้อมูลในการวิจัยครั้งนี้ได้

5.1.1.1 สรุปผลประสิทธิภาพการวิเคราะห์กลุ่มข้อมูลที่จำลองขึ้น

จากการวิจัย พบว่า ระยะห่างที่ทำให้การวิเคราะห์กลุ่มสำหรับข้อมูลกรณีต่าง ๆ มีประสิทธิภาพดีที่สุดโดยเฉลี่ย แตกต่างกันไปตามรูปแบบของข้อมูลนั้น ๆ ดังตารางที่ 5.1 อย่างไรก็ตาม แม้ว่าระยะห่างที่ทำให้การวิเคราะห์กลุ่มมีประสิทธิภาพดีที่สุดโดยเฉลี่ย อาจไม่แตกต่างจากการวิเคราะห์กลุ่มด้วยระยะห่างแบบอื่น ๆ ที่ระดับนัยสำคัญ 0.05

ตารางที่ 5.2 ระยะห่างที่ทำให้การวิเคราะห์กลุ่มสำหรับข้อมูลกรณีต่าง ๆ มีประสิทธิภาพดีที่สุดโดยเฉลี่ย

K	ρ	n	ข้อมูลรูปแบบที่					
			I	II	III	IV	V	VI
3	0.2	20	KR&N [1]	P	KR&N [1]	KR [4]	KR [4]	P
		100	KR&N [1]	P[3]	KR&N [1]	KR [4]	KR [4]	KR , P
	0.8	20	KR&N [2]	KR	KR&N [1]	KR [4]	KR [4]	P
		100	KR&N [1]	KR	KR&N [1]	KR [4]	KR [4]	KR , P
5	0.2	20	P&N	KR	KR	P	KR [4]	P
		100	KR&N , P&N	KR	KR&N [1]	P	KR [4]	KR , P
	0.8	20	KR&N , P&N	P	KR&N	P	KR	KR , P
		100	P&N	P&N	KR&N	P	P	KR , P

หมายเหตุ

[1] ประสิทธิภาพการวิเคราะห์กลุ่มด้วยระยะห่างแบบ KR&N แตกต่างจากการวิเคราะห์กลุ่มด้วยระยะห่างแบบอื่น ๆ ที่ระดับนัยสำคัญ 0.05

[2] ประสิทธิภาพการวิเคราะห์กลุ่มด้วยระยะห่างแบบ KR&N แตกต่างจากการวิเคราะห์กลุ่มด้วยระยะห่างแบบอื่น ๆ ที่ระดับนัยสำคัญ 0.05 ยกเว้นค่าความบริสุทธิ์

[3] ประสิทธิภาพการวิเคราะห์กลุ่มด้วยระยะห่างของ P แตกต่างจากการวิเคราะห์กลุ่มด้วยระยะห่างแบบอื่น ๆ ที่ระดับนัยสำคัญ 0.05

[4] ประสิทธิภาพการวิเคราะห์กลุ่มด้วยระยะห่างของ KR แตกต่างจากการวิเคราะห์กลุ่มด้วยระยะห่างแบบอื่น ๆ ที่ระดับนัยสำคัญ 0.05

พิจารณาประสิทธิภาพในการวิเคราะห์กลุ่มในภาพรวม โดยมุ่งประเด็นไปที่ขอบเขตการศึกษาที่มีอิทธิพลต่อประสิทธิภาพการวิเคราะห์กลุ่ม ได้แก่ รูปแบบของข้อมูล จำนวนกลุ่ม ระดับความสัมพันธ์หรือค่าสัมประสิทธิ์สหสัมพันธ์ และขนาดข้อมูลต่อกลุ่ม ซึ่งสรุปผลการวิเคราะห์กลุ่มได้ดังนี้

กรณีจำนวนกลุ่มเท่ากับ 3 (K = 3)

ในการวิจัยครั้งนี้ พิจารณากรณีข้อมูลถูกแบ่งกลุ่มเป็น 3 กลุ่ม เพื่อเป็นตัวแทนของข้อมูลที่มีจำนวนกลุ่ม น้อยกว่าจำนวนประเภทหรืออันดับของตัวแปรนามบัญญัติและตัวแปรอันดับ ซึ่งมีค่าเท่ากับ 5 และเป็นตัวแทนของข้อมูลที่มีจำนวนความถี่ของแต่ละประเภทหรืออันดับแตกต่างกัน

- *ข้อมูลรูปแบบที่ I*

สำหรับข้อมูลรูปแบบที่ I ซึ่งเป็นข้อมูลแบบผสมที่ประกอบไปด้วยตัวแปรทั้ง 3 ชนิด การวิเคราะห์กลุ่มด้วยระยะห่างแบบ KR&N มีประสิทธิภาพดีที่สุดโดยเฉลี่ย และแตกต่างจากการวิเคราะห์กลุ่มด้วยระยะห่างแบบอื่น ๆ ที่ระดับนัยสำคัญ 0.05 แสดงว่าระยะห่างสำหรับตัวแปรเชิงปริมาณของ Gower ระยะห่างสำหรับตัวแปรอันดับของ Kaufman and Rousseeuw และระยะห่างสำหรับตัวแปรนามบัญญัติของ Noorbehbahani et al. ที่นำมาใช้ร่วมกัน เหมาะสำหรับข้อมูลแบบผสมนี้ ขณะที่การวิเคราะห์กลุ่มด้วยระยะห่างแบบ P&N มีประสิทธิภาพด้อยกว่า แม้ว่าจะวัดระยะห่างสำหรับตัวแปรนามบัญญัติด้วยวิธีเดียวกัน แสดงว่าการวัดระยะห่างสำหรับตัวแปรนามบัญญัติของ Noorbehbahani et al. ร่วมกับระยะห่างสำหรับตัวแปรอันดับของ Kaufman and Rousseeuw มีประสิทธิภาพดีกว่าการวัดระยะห่างสำหรับตัวแปรนามบัญญัติของ Noorbehbahani et al. ร่วมกับระยะห่างสำหรับตัวแปรอันดับของ Podani

พิจารณาตามระดับความสัมพันธ์

เมื่อค่าสัมประสิทธิ์สหสัมพันธ์เพิ่มขึ้นจาก 0.2 เป็น 0.8 โดยเฉลี่ยการวิเคราะห์กลุ่มด้วยระยะห่างทั้ง 4 วิธีมีประสิทธิภาพลดลง (ค่า Purity ค่า Rand statistic ค่า Jaccard coefficient และค่า Average silhouette width ลดลง) นอกจากนี้การวิเคราะห์กลุ่มด้วยระยะห่างของ KR มีประสิทธิภาพขึ้นมาอยู่อันดับที่ 2 แสดงว่าเมื่อตัวแปรมีความสัมพันธ์กันมากขึ้น การวัดระยะห่างสำหรับตัวแปรอันดับของ Podani ทำให้ประสิทธิภาพการวิเคราะห์กลุ่มลดลง

พิจารณาตามขนาดข้อมูลต่อกลุ่ม

เมื่อขนาดข้อมูลต่อกลุ่มเพิ่มขึ้นจาก 20 เป็น 100 ช่วงความเชื่อมั่น 95% ของค่าต่าง ๆ ในการวัดประสิทธิภาพการวิเคราะห์กลุ่มลดลง แต่อันดับประสิทธิภาพการวิเคราะห์กลุ่มยังคงเดิม

ตารางที่ 5.3 อันดับประสิทธิภาพการวิเคราะห์กลุ่มโดยเฉลี่ย ข้อมูลรูปแบบที่ I เมื่อ $K = 3$

ข้อมูลรูปแบบที่ I			อันดับที่			
K	ρ	n	1	2	3	4
3	0.2	20	KR&N	P&N	P	KR
		100	KR&N	P&N	P	KR
	0.8	20	KR&N	KR	P&N	P
		100	KR&N	KR	P&N	P

- *ข้อมูลรูปแบบที่ II*

สำหรับข้อมูลรูปแบบที่ II ซึ่งเป็นข้อมูลที่ไม่มีตัวแปรเชิงปริมาณ แต่ประกอบไปด้วยตัวแปรนามบัญญัติและตัวแปรอันดับเท่านั้น การวิเคราะห์กลุ่มด้วยระยะห่างแบบ KR&N และระยะห่างแบบ P&N มีประสิทธิภาพอยู่ใน 2 อันดับสุดท้าย ขณะที่การวิเคราะห์กลุ่มด้วยระยะห่างของ KR และระยะห่างของ P มีประสิทธิภาพอยู่ใน 2 อันดับแรก ซึ่งแตกต่างจากการวิเคราะห์กลุ่มข้อมูลรูปแบบที่ I แสดงว่าการวัดระยะห่างสำหรับตัวแปรนามบัญญัติของ Noorbehbahani et al. อาจไม่เหมาะสมสำหรับการวิเคราะห์กลุ่มข้อมูลที่ไม่ประกอบไปด้วยตัวแปรเชิงปริมาณ

พิจารณาตามระดับความสัมพันธ์

เมื่อค่าสัมประสิทธิ์สหสัมพันธ์เพิ่มขึ้นจาก 0.2 เป็น 0.8 การวิเคราะห์กลุ่มด้วยระยะห่างทั้ง 4 วิธีมีประสิทธิภาพลดลง (ค่า Purity ค่า Rand statistic ค่า Jaccard coefficient และค่า Average silhouette width ลดลง) นอกจากนี้การวิเคราะห์กลุ่มด้วยระยะห่างของ KR มีประสิทธิภาพขึ้นเป็นอันดับ 1 แทนระยะห่างของ P แต่การวิเคราะห์กลุ่มด้วยระยะห่างทั้ง 2 นี้มีประสิทธิภาพไม่แตกต่างกัน

พิจารณาตามขนาดข้อมูลต่อกลุ่ม

เมื่อขนาดข้อมูลต่อกลุ่มเพิ่มขึ้นจาก 20 เป็น 100 ช่วงความเชื่อมั่น 95% ของค่าต่าง ๆ ในการวัดประสิทธิภาพการวิเคราะห์กลุ่มลดลง และอันดับประสิทธิภาพการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ ยังคงเดิม

ตารางที่ 5.4 อันดับประสิทธิภาพการวิเคราะห์กลุ่มโดยเฉลี่ย ข้อมูลรูปแบบที่ II เมื่อ K = 3

ข้อมูลรูปแบบที่ II			อันดับที่			
K	ρ	n	1	2	3	4
3	0.2	20	P	KR	KR&N	P&N
		100	P	KR	KR&N	P&N
	0.8	20	KR	P	KR&N	P&N
		100	KR	P	KR&N	P&N

- *ข้อมูลรูปแบบที่ III*

สำหรับข้อมูลรูปแบบที่ III ซึ่งเป็นข้อมูลที่ประกอบไปด้วยตัวแปรนามบัญญัติและตัวแปรเชิงปริมาณเท่านั้น การวิเคราะห์กลุ่มด้วยระยะห่างแบบ KR&N มีประสิทธิภาพดีกว่าการวิเคราะห์กลุ่มด้วยระยะห่างของ KR และแตกต่างกันที่ระดับนัยสำคัญ 0.05 แสดงว่าการวัดระยะห่างสำหรับตัวแปรเชิงปริมาณของ Gower ร่วมกับระยะห่างสำหรับตัวแปรนามบัญญัติของ Noorbehbahani et al. ให้ผลการวิเคราะห์กลุ่มที่มีประสิทธิภาพดีกว่าระยะห่างสำหรับตัวแปรเชิงปริมาณและตัวแปรนามบัญญัติของ Gower สำหรับข้อมูลรูปแบบนี้

พิจารณาตามระดับความสัมพันธ์

เมื่อค่าสัมประสิทธิ์สหสัมพันธ์เพิ่มขึ้นจาก 0.2 เป็น 0.8 การวิเคราะห์กลุ่มด้วยระยะห่างทั้ง 2 วิธีมีประสิทธิภาพลดลง (ค่า Purity ค่า Rand statistic ค่า Jaccard coefficient และค่า Average silhouette width ลดลง) และอันดับประสิทธิภาพการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ ยังคงเดิม

พิจารณาตามขนาดข้อมูลต่อกลุ่ม

เมื่อขนาดข้อมูลต่อกลุ่มเพิ่มขึ้นจาก 20 เป็น 100 ช่วงความเชื่อมั่น 95% ของค่าต่าง ๆ ในการวัดประสิทธิภาพการวิเคราะห์กลุ่มลดลง และอันดับประสิทธิภาพการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ ยังคงเดิม

ตารางที่ 5.5 อันดับประสิทธิภาพการวิเคราะห์กลุ่มโดยเฉลี่ย ข้อมูลรูปแบบที่ III เมื่อ K = 3

ข้อมูลรูปแบบที่ III			อันดับที่	
K	ρ	n	1	2
3	0.2	20	KR&N	KR
		100	KR&N	KR
	0.8	20	KR&N	KR
		100	KR&N	KR

- *ข้อมูลรูปแบบที่ IV*

สำหรับข้อมูลรูปแบบที่ IV ซึ่งเป็นข้อมูลที่ประกอบไปด้วยตัวแปรอันดับและตัวแปรเชิงปริมาณเท่านั้น การวิเคราะห์กลุ่มด้วยระยะห่างของ KR มีประสิทธิภาพดีกว่าการวิเคราะห์กลุ่มด้วยระยะห่างของ P และแตกต่างกันที่ระดับนัยสำคัญ 0.05 แสดงว่าการวัดระยะห่างสำหรับตัวแปรเชิงปริมาณของ Gower ร่วมกับตัวแปรอันดับของ Kaufman and Rousseeuw ให้ผลการวิเคราะห์กลุ่มที่มีประสิทธิภาพดีกว่าระยะห่างสำหรับตัวแปรเชิงปริมาณของ Gower ร่วมกับตัวแปรอันดับของ Podani สำหรับข้อมูลรูปแบบนี้

พิจารณาตามระดับความสัมพันธ์

เมื่อค่าสัมประสิทธิ์สหสัมพันธ์เพิ่มขึ้นจาก 0.2 เป็น 0.8 การวิเคราะห์กลุ่มด้วยระยะห่างทั้ง 2 วิธีมีประสิทธิภาพลดลง (ค่า Purity ค่า Rand statistic ค่า Jaccard coefficient และค่า Average silhouette width ลดลง) และอันดับประสิทธิภาพการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ ยังคงเดิม

พิจารณาตามขนาดข้อมูลต่อกลุ่ม

เมื่อขนาดข้อมูลต่อกลุ่มเพิ่มขึ้นจาก 20 เป็น 100 อันดับประสิทธิภาพการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ ยังคงเดิม

ตารางที่ 5.6 อันดับประสิทธิภาพการวิเคราะห์กลุ่มโดยเฉลี่ย ข้อมูลรูปแบบที่ IV เมื่อ K = 3

ข้อมูลรูปแบบที่ IV			อันดับที่	
K	ρ	n	1	2
3	0.2	20	KR	P
		100	KR	P
	0.8	20	KR	P
		100	KR	P

- *ข้อมูลรูปแบบที่ V*

สำหรับข้อมูลรูปแบบที่ V ซึ่งเป็นข้อมูลที่ประกอบไปด้วยตัวแปรนามบัญญัติเท่านั้น การวิเคราะห์กลุ่มด้วยระยะห่างของ KR มีประสิทธิภาพดีกว่าการวิเคราะห์กลุ่มด้วยระยะห่างแบบ KR&N และแตกต่างกันที่ระดับนัยสำคัญ 0.05 แสดงว่าในกรณีที่ข้อมูลไม่มีตัวแปรเชิงปริมาณและตัวแปรอันดับ ระยะห่างสำหรับตัวแปรเชิงปริมาณของ Gower ให้ผลการวิเคราะห์กลุ่มที่มีประสิทธิภาพดีกว่าระยะห่างสำหรับตัวแปรนามบัญญัติของ Noorbehbahani et al. สำหรับข้อมูลรูปแบบนี้

พิจารณาตามระดับความสัมพันธ์

เมื่อค่าสัมประสิทธิ์สหสัมพันธ์เพิ่มขึ้นจาก 0.2 เป็น 0.8 การวิเคราะห์กลุ่มด้วยระยะห่างทั้ง 2 วิธีมีประสิทธิภาพลดลง (ค่า Purity ค่า Rand statistic ค่า Jaccard coefficient และค่า Average silhouette width ลดลง) และอันดับประสิทธิภาพการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ ยังคงเดิม

พิจารณาตามขนาดข้อมูลต่อกลุ่ม

เมื่อขนาดข้อมูลต่อกลุ่มเพิ่มขึ้นจาก 20 เป็น 100 อันดับประสิทธิภาพการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ ยังคงเดิม

ตารางที่ 5.7 อันดับประสิทธิภาพการวิเคราะห์กลุ่มโดยเฉลี่ย ข้อมูลรูปแบบที่ V เมื่อ $K = 3$

ข้อมูลรูปแบบที่ V			อันดับที่	
K	ρ	n	1	2
3	0.2	20	KR	KR&N
		100	KR	KR&N
	0.8	20	KR	KR&N
		100	KR	KR&N

- *ข้อมูลรูปแบบที่ VI*

สำหรับข้อมูลรูปแบบที่ VI ซึ่งเป็นข้อมูลที่ประกอบไปด้วยตัวแปรอันดับเพียงชนิดเดียว การวิเคราะห์กลุ่มด้วยระยะห่างของ KR มีประสิทธิภาพไม่แตกต่างจากการวิเคราะห์กลุ่มด้วยระยะห่างของ P แสดงว่าการวัดระยะห่างสำหรับตัวแปรอันดับของ Kaufman and Rousseeuw ให้ผลการวิเคราะห์กลุ่มที่มีประสิทธิภาพเช่นเดียวกับระยะห่างสำหรับตัวแปรอันดับของ Podani สำหรับข้อมูลรูปแบบนี้

พิจารณาตามระดับความสัมพันธ์

เมื่อค่าสัมประสิทธิ์สหสัมพันธ์เพิ่มขึ้นจาก 0.2 เป็น 0.8 การวิเคราะห์กลุ่มด้วยระยะห่างทั้ง 2 วิธีมีประสิทธิภาพลดลง (ค่า Purity ค่า Rand statistic ค่า Jaccard coefficient และค่า Average silhouette width ลดลง) และอันดับประสิทธิภาพการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ ยังคงเดิม

พิจารณาตามขนาดข้อมูลต่อกลุ่ม

เมื่อขนาดข้อมูลต่อกลุ่มเพิ่มขึ้นจาก 20 เป็น 100 อันดับประสิทธิภาพการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ ยังคงเดิม

ตารางที่ 5.8 อันดับประสิทธิภาพการวิเคราะห์กลุ่มโดยเฉลี่ย ข้อมูลรูปแบบที่ VI เมื่อ K = 3

ข้อมูลรูปแบบที่ VI			อันดับที่	
K	ρ	n	1	2
3	0.2	20	P	KR
		100	KR	P
	0.8	20	P	KR
		100	KR	P

กรณีจำนวนกลุ่มเท่ากับ 5 (K = 5)

ในการวิจัยครั้งนี้ พิจารณากรณีข้อมูลถูกแบ่งกลุ่มเป็น 5 กลุ่ม เพื่อเป็นตัวแทนของข้อมูลที่มีจำนวนกลุ่ม เท่ากับจำนวนประเภทหรืออันดับของตัวแปรนามบัญญัติและตัวแปรอันดับ ซึ่งมีค่าเท่ากับ 5 และเป็นตัวแทนของข้อมูลที่มีจำนวนความถี่ของแต่ละประเภทหรืออันดับเท่ากัน ผลการวิจัยพบว่า สำหรับข้อมูลทั้ง 6 รูปแบบ การวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ มีประสิทธิภาพไม่แตกต่างกัน เมื่อทดสอบความแตกต่างในระดับนัยสำคัญทางสถิติ ยกเว้น 3 กรณีดังต่อไปนี้

กรณีข้อมูลรูปแบบที่ III เมื่อ K = 5, ρ = 0.2 และ n = 100

กรณีข้อมูลรูปแบบที่ IV เมื่อ K = 5, ρ = 0.2 และ n = 20

กรณีข้อมูลรูปแบบที่ IV เมื่อ K = 5, ρ = 0.2 และ n = 100

ซึ่งการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ มีประสิทธิภาพแตกต่างกัน ทั้งนี้ข้อมูลทั้ง 3 กรณี มีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากัน คือ ρ = 0.2 หรือตัวแปรมีความสัมพันธ์กันน้อย และเป็นข้อมูลที่มีตัวแปรนามบัญญัติเป็นส่วนประกอบ นั่นคือมีการวัดระยะห่างสำหรับตัวแปรนามบัญญัติทั้งของ Gower และของ Noorbehbahani et al.

5.1.1.2 สรุปผลค่า Average silhouette width จากการวิเคราะห์กลุ่มข้อมูลที่จำลองขึ้น

เนื่องจากค่า Average Silhouette Width เป็นค่าที่วัดประสิทธิภาพการวิเคราะห์กลุ่ม โดยเปรียบเทียบค่า Average Silhouette Width ที่ได้จากการวิเคราะห์กลุ่มที่กำหนดจำนวนกลุ่มแตกต่างกัน เพื่อหาจำนวนกลุ่มที่เหมาะสมที่สุดสำหรับข้อมูลชุดนั้น ๆ เมื่อไม่ทราบจำนวนกลุ่มที่แท้จริงหรือจำนวนกลุ่มที่ต้องการ และค่า Average Silhouette Width นี้ต้องคำนวณจากระยะห่างระหว่างข้อมูล ดังนั้นในการวิจัยนี้จึงไม่นำค่า Average Silhouette Width มาเปรียบเทียบประสิทธิภาพการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ เนื่องจากระยะห่างแต่ละวิธีต่างให้ความแตกต่างระหว่างกลุ่มในระดับทศนิยมที่แตกต่างกัน แต่ได้คำนวณค่า Average Silhouette Width เพื่อสังเกตทิศทางของค่า Average Silhouette Width จากการวิเคราะห์กลุ่มด้วยระยะห่างแบบต่าง ๆ เท่านั้น

สำหรับข้อมูลกรณี K = 3 ข้อมูลรูปแบบเดียวกันจะให้ค่าเฉลี่ยของค่า Average silhouette width จากการวิเคราะห์กลุ่มข้อมูลด้วยระยะห่างแบบต่าง ๆ ที่อันดับเดียวกัน ดังตารางที่ 5.8 และอันดับต่าง ๆ แตกต่างกันที่ระดับนัยสำคัญ 0.05 ทุกกรณี ขณะที่ข้อมูลกรณี K = 5 ค่า Average silhouette width จากการวิเคราะห์กลุ่มข้อมูลด้วยระยะห่างแบบต่าง ๆ มีอันดับดังตารางที่ 5.9 ซึ่งมีทั้งกรณีที่อันดับแตกต่างกันที่ระดับนัยสำคัญ 0.05 และไม่แตกต่างกัน

ตารางที่ 5.9 อันดับค่า Average silhouette width จากการวิเคราะห์กลุ่มข้อมูลกรณี K = 3 ด้วยระยะห่างแบบต่าง ๆ

ข้อมูลรูปแบบที่	K = 3		อันดับที่			
	ρ	n	1	2	3	4
I	0.2	20	P	KR	P&N	KR&N
		100	P	KR	P&N	KR&N
	0.8	20	P	KR	P&N	KR&N
		100	P	KR	P&N	KR&N
II	0.2	20	P	KR	P&N	KR&N
		100	P	KR	P&N	KR&N
	0.8	20	P	KR	P&N	KR&N
		100	P	KR	P&N	KR&N
III	0.2	20	KR	KR&N		
		100	KR	KR&N		
	0.8	20	KR	KR&N		
		100	KR	KR&N		
IV	0.2	20	P	KR		
		100	P	KR		
	0.8	20	P	KR		
		100	P	KR		
V	0.2	20	KR	KR&N		
		100	KR	KR&N		
	0.8	20	KR	KR&N		
		100	KR	KR&N		
VI	0.2	20	P	KR		
		100	P	KR		
	0.8	20	P	KR		
		100	P	KR		

ตารางที่ 5.10 อันดับค่า Average silhouette width จากการวิเคราะห์กลุ่มข้อมูลกรณี K = 5 ด้วยระยะห่างแบบต่าง ๆ

ข้อมูลรูปแบบที่	K = 5		อันดับที่			
	ρ	n	1	2	3	4
I	0.2	20	P	KR	P&N	KR&N
		100	P	KR	P&N	KR&N
	0.8	20	P	KR	P&N	KR&N
		100	P	KR	P&N	KR&N
II	0.2	20	P	KR	P&N	KR&N
		100	KR	P	P&N	KR&N
	0.8	20	KR	P	P&N	KR&N
		100	KR	P	P&N	KR&N
III	0.2	20	KR&N	KR		
		100	KR&N	KR		
	0.8	20	KR	KR&N		
		100	KR	KR&N		
IV	0.2	20	P	KR		
		100	P	KR		
	0.8	20	P	KR		
		100	P	KR		
V	0.2	20	KR&N	KR		
		100	KR&N	KR		
	0.8	20	KR&N	KR		
		100	KR&N	KR		
VI	0.2	20	P	KR		
		100	P	KR		
	0.8	20	KR	P		
		100	KR	P		

5.1.2 สรุปผลการวิเคราะห์กลุ่มข้อมูลจริง

ข้อมูล Pittsburgh Bridges Data Set เป็นตัวอย่างข้อมูลจริงสำหรับศึกษาการวิเคราะห์กลุ่มข้อมูลที่ทราบกลุ่มที่แท้จริง ซึ่งเป็นข้อมูลแบบผสม โดยมีตัวแปรนามบัญญัติ จำนวน 7 ตัวแปร ตัวแปรอันดับ จำนวน 1 ตัวแปร และตัวแปรเชิงปริมาณ จำนวน 3 ตัวแปร

ผลการวิเคราะห์กลุ่มด้วยระยะห่างทั้ง 4 วิธี พบว่า การวิเคราะห์กลุ่มด้วยระยะห่างแบบ KR&N ให้ค่า Purity และค่า Jaccard coefficient สูงที่สุด แต่กลับให้ค่า Rand statistic ต่ำที่สุด แสดงว่าการวิเคราะห์กลุ่มด้วยระยะห่างแบบ KR&N มีความถูกต้องในภาพรวมและสามารถจัดข้อมูลคู่ที่ทราบว่าอยู่กลุ่มเดียวกันให้อยู่กลุ่มเดียวกันมากที่สุด แต่สำหรับข้อมูลคู่ที่ทราบว่าอยู่ต่างกลุ่มกันกลับถูกจัดให้อยู่ต่างกลุ่มกันน้อยกว่าวิธีอื่น ๆ โดยเฉพาะการวิเคราะห์กลุ่มด้วยระยะห่างแบบ P&N ที่ให้ผลดีกว่า ซึ่งให้ค่า Rand statistic สูงที่สุด

อย่างไรก็ตามค่า Jaccard coefficient จากการวิเคราะห์กลุ่มทั้ง 4 วิธี มีค่าประมาณ 0.2 ซึ่งใกล้ 0 แสดงว่าการวิเคราะห์กลุ่มให้ผลถูกต้องใกล้เคียงกับกลุ่มที่แท้จริงไม่มากนัก ขณะที่ค่า Purity และค่า Rand statistic จากการวิเคราะห์กลุ่มด้วยระยะห่างทั้ง 4 วิธี มีค่าอยู่ระหว่าง 0.46 ถึง 0.52 และ 0.74 ถึง 0.77 ตามลำดับ ซึ่งเป็นระดับที่ยอมรับได้ นอกจากนี้ค่า Average silhouette width มีค่าอยู่ระหว่าง 0.23 ถึง 0.29 ซึ่งเป็นไปได้ว่าการปรับจำนวนกลุ่มข้อมูลอาจทำให้ผลการวิเคราะห์กลุ่มดีขึ้น

5.2 อภิปรายผลการวิจัย

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพในการวิเคราะห์กลุ่มด้วยอัลกอริทึมจัดกลุ่มโดยรอบมีดอยด์ ที่ใช้ระยะห่างของ KR ระยะห่างของ P ระยะห่างแบบ KR&N และระยะห่างแบบ P&N สำหรับข้อมูลที่ประกอบไปด้วยตัวแปรนามบัญญัติ ตัวแปรอันดับ และตัวแปรเชิงปริมาณ โดยเฉพาะข้อมูลแบบผสมที่ประกอบไปด้วยตัวแปรทั้ง 3 ชนิดนี้ โดยทำการจำลองข้อมูลที่กำหนดให้ทราบกลุ่มแน่ชัด มีการพิจารณากรณีที่ตัวแปรมีความสัมพันธ์กันน้อยและมีความสัมพันธ์กันมาก ขนาดข้อมูลต่อกลุ่มที่แตกต่างกัน ทั้งยังพิจารณากรณีที่จำนวนความถี่ของแต่ละประเภทหรืออันดับแตกต่างกัน และไม่แตกต่างกัน ด้วยการกำหนดจำนวนกลุ่มของข้อมูลที่แตกต่างกัน ศึกษาประสิทธิภาพการวิเคราะห์กลุ่มโดยเปรียบเทียบค่า Purity ค่า Rand statistic และค่า Jaccard coefficient พบว่าทั้งสามค่าให้ผลการวิเคราะห์กลุ่มเป็นไปในทิศทางเดียวกัน ขณะที่ค่า Average silhouette width ให้ผลที่แตกต่างออกไป

การวิเคราะห์กลุ่มข้อมูล เมื่อข้อมูลถูกแบ่งกลุ่มเป็น 3 กลุ่ม นั่นคือมีจำนวนกลุ่มน้อยกว่าจำนวนประเภทและจำนวนอันดับ ซึ่งเป็นกรณีข้อมูลที่จำนวนความถี่ของข้อมูลแต่ละประเภทหรือ

อันดับแตกต่างกัน พบว่า โดยส่วนใหญ่การวิเคราะห์กลุ่มด้วยระยะห่างทั้ง 4 วิธีแตกต่างกันอย่างมีนัยสำคัญ สำหรับข้อมูลรูปแบบที่ I ซึ่งเป็นข้อมูลแบบผสมที่ประกอบไปด้วยตัวแปรทั้ง 3 ชนิด การวิเคราะห์กลุ่มด้วยระยะห่างแบบ KR&N มีประสิทธิภาพดีที่สุดโดยเฉลี่ย และแตกต่างจากการวิเคราะห์กลุ่มด้วยระยะห่างแบบอื่น ๆ อย่างมีนัยสำคัญ รองลงมาคือการวิเคราะห์กลุ่มด้วยระยะห่างแบบ P&N ขณะที่สำหรับข้อมูลรูปแบบที่ II ซึ่งเป็นข้อมูลที่ไม่มีตัวแปรเชิงปริมาณ แต่ประกอบไปด้วยตัวแปรนามบัญญัติและตัวแปรอันดับเท่านั้น การวิเคราะห์กลุ่มด้วยระยะห่างแบบ KR&N และระยะห่างแบบ P&N มีประสิทธิภาพอยู่ใน 2 อันดับสุดท้าย แต่การวิเคราะห์กลุ่มด้วยระยะห่างของ KR และระยะห่างของ P มีประสิทธิภาพอยู่ใน 2 อันดับแรก เป็นที่สังเกตว่าการวัดระยะห่างสำหรับตัวแปรนามบัญญัติของ Noorbehbahani et al. อาจไม่เหมาะสมสำหรับการวิเคราะห์กลุ่มข้อมูลที่ไม่ประกอบไปด้วยตัวแปรเชิงปริมาณ

พิจารณาข้อมูลรูปแบบที่ III ซึ่งเป็นข้อมูลที่ประกอบไปด้วยตัวแปรนามบัญญัติและตัวแปรเชิงปริมาณเท่านั้น การวิเคราะห์กลุ่มด้วยระยะห่างแบบ KR&N มีประสิทธิภาพดีกว่าการวิเคราะห์กลุ่มด้วยระยะห่างของ KR และแตกต่างกันอย่างมีนัยสำคัญ แต่เมื่อพิจารณาการวิเคราะห์กลุ่มข้อมูลรูปแบบที่ V ซึ่งเป็นข้อมูลที่มีตัวแปรนามบัญญัติเพียงชนิดเดียว การวิเคราะห์กลุ่มด้วยระยะห่างของ KR กลับมีประสิทธิภาพดีกว่าการวิเคราะห์กลุ่มด้วยระยะห่างแบบ KR&N และแตกต่างกันอย่างมีนัยสำคัญ

ดังนั้นจากผลการวิเคราะห์กลุ่มข้อมูลรูปแบบที่ I II III และ V จึงอาจสรุปได้ว่าในกรณีที่จำนวนกลุ่มน้อยกว่าจำนวนประเภทและจำนวนอันดับ ทำให้ในหนึ่งกลุ่มข้อมูลอาจมีข้อมูลตัวแปรนั้นมากกว่าหนึ่งประเภท ระยะห่างแบบ KR&N เหมาะสำหรับการวิเคราะห์กลุ่มข้อมูลที่ประกอบไปด้วยทั้งตัวแปรเชิงปริมาณและตัวแปรนามบัญญัติ และอาจมีหรือไม่มีตัวแปรอันดับก็ได้ ทั้งนี้อาจเป็นเพราะว่าการวัดระยะห่างสำหรับตัวแปรนามบัญญัติของ Noorbehbahani et al. ทำให้ระยะห่างระหว่างข้อมูลมีความละเอียดมากกว่าการวัดระยะห่างสำหรับตัวแปรนามบัญญัติของ Gower ขณะที่การวัดระยะห่างสำหรับตัวแปรเชิงปริมาณของ Gower แบ่งกลุ่มข้อมูลตัวแปรเชิงปริมาณเท่านั้น เมื่อใช้ร่วมกับการวัดระยะห่างสำหรับตัวแปรนามบัญญัติของ Noorbehbahani et al. ซึ่งสามารถแบ่งแยกความแตกต่างระหว่างข้อมูลตัวแปรนามบัญญัตินั้น ๆ ที่อยู่ต่างประเภทกันได้ดีกว่า จึงทำให้การวิเคราะห์ข้อมูลมีประสิทธิภาพดีกว่าระยะห่างแบบอื่น ๆ อย่างไรก็ตามเมื่อข้อมูลแบบผสมที่ประกอบไปด้วยตัวแปรทั้ง 3 ชนิด การวิเคราะห์กลุ่มด้วยระยะห่างแบบ P&N มีประสิทธิภาพรองลงมา แม้ว่าจะวัดระยะห่างสำหรับตัวแปรนามบัญญัติด้วยวิธีของ Noorbehbahani et al. เช่นเดียวกัน อาจเป็นเพราะความแตกต่างของจำนวนความถี่ของอันดับของตัวแปรอันดับ ทำให้เมื่อวัดระยะห่างสำหรับตัวแปรอันดับตามนิยามของ Podani เกิดระยะห่างระหว่างข้อมูลอันดับที่มีความถี่สูงกับข้อมูลอันดับที่มีความถี่ต่ำมากเกินไป ทำให้การวัดระยะห่างสำหรับตัวแปรอันดับของ

Podani ร่วมกับระยะห่างสำหรับตัวแปรนามบัญญัติของ Noorbehbahani et al. มีความผิดพลาดมากกว่าการวัดระยะห่างสำหรับตัวแปรอันดับของ Kaufman and Rousseeuw ร่วมกับระยะห่างสำหรับตัวแปรนามบัญญัติของ Noorbehbahani et al.

สำหรับข้อมูลรูปแบบที่ IV ซึ่งเป็นข้อมูลที่ประกอบไปด้วยตัวแปรอันดับและตัวแปรเชิงปริมาณเท่านั้น การวิเคราะห์กลุ่มด้วยระยะห่างของ KR มีประสิทธิภาพดีกว่าการวิเคราะห์กลุ่มด้วยระยะห่างของ P และแตกต่างกันอย่างมีนัยสำคัญ ขณะที่สำหรับข้อมูลรูปแบบที่ VI ซึ่งเป็นข้อมูลที่ประกอบไปด้วยตัวแปรอันดับเพียงชนิดเดียว การวิเคราะห์กลุ่มด้วยระยะห่างของ KR มีประสิทธิภาพไม่แตกต่างจากการวิเคราะห์กลุ่มด้วยระยะห่างของ P จึงอาจสรุปได้ว่าการวัดระยะห่างสำหรับตัวแปรอันดับของ Kaufman and Rousseeuw ส่งผลให้การวิเคราะห์กลุ่มมีประสิทธิภาพดีกว่าการวัดระยะห่างสำหรับตัวแปรอันดับของ Podani สำหรับข้อมูลที่ประกอบไปด้วยทั้งตัวแปรอันดับและตัวแปรเชิงปริมาณเท่านั้น

เมื่อข้อมูลถูกแบ่งกลุ่มเป็น 5 กลุ่ม นั่นคือมีจำนวนกลุ่มเท่ากับจำนวนประเภทหรืออันดับ ซึ่งเป็นกรณีข้อมูลที่มีจำนวนความถี่ของแต่ละประเภทหรืออันดับไม่แตกต่างกัน พบว่า โดยส่วนใหญ่การวิเคราะห์กลุ่มด้วยระยะห่างทั้ง 4 วิธีไม่แตกต่างกัน นั่นคือการวิเคราะห์กลุ่มด้วยระยะห่างของ KR ระยะห่างของ P ระยะห่างแบบ KR&N และระยะห่างแบบ P&N มีประสิทธิภาพไม่แตกต่างกัน อาจเป็นเพราะว่าการวัดระยะห่างสำหรับตัวแปรนามบัญญัติของ Noorbehbahani et al. และการวัดระยะห่างสำหรับตัวแปรอันดับของ Podani ซึ่งมีค่าขึ้นกับจำนวนความถี่ของประเภทหรืออันดับ เมื่อจำนวนความถี่ของแต่ละประเภทหรืออันดับไม่แตกต่างกัน การวัดระยะห่างสำหรับตัวแปรนามบัญญัติของ Noorbehbahani et al. จึงไม่แตกต่างจากของ Gower และการวัดระยะห่างสำหรับตัวแปรอันดับของ Podani จึงไม่แตกต่างจากของ Kaufman and Rousseeuw

เนื่องจากการศึกษาวิจัยนี้ไม่สามารถจำลองข้อมูลแบบผสมครอบคลุมทุกกรณีที่เป็นไปได้ เพราะว่าในความเป็นจริงข้อมูลมีความหลากหลายสูง ทั้งด้านจำนวนตัวแปรแต่ละชนิดที่แตกต่างกัน จำนวนประเภทหรืออันดับที่มีค่าที่เป็นไปได้มากมาย จำนวนกลุ่มของข้อมูล และความสัมพันธ์ระหว่างตัวแปรที่เป็นไปได้หลายค่า การวัดประสิทธิภาพในการวิเคราะห์กลุ่มด้วยอัลกอริทึมจัดกลุ่มโดยรอบมีดอยด์ที่ใช้มาตรวัดระยะห่างแบบต่าง ๆ ในการศึกษาวิจัยนี้ จึงสามารถสรุปได้แน่ชัดในบางกรณีเท่านั้น และอาจเป็นแนวทางการศึกษาข้อมูลแบบผสมกรณีอื่น ๆ ต่อไป อย่างไรก็ตามระยะห่างแบบ KR&N และระยะห่างแบบ P&N ที่นำเสนอขึ้นใหม่ ให้ผลการวิเคราะห์กลุ่มสำหรับข้อมูลแบบผสมที่ดีในกรณีที่จำนวนความถี่ของแต่ละประเภทหรืออันดับของข้อมูลแตกต่างกัน ซึ่งเป็นลักษณะทั่วไปของข้อมูลจริง ที่ไม่สามารถกำหนดหรือควบคุมได้ว่าข้อมูลตัวแปรนามบัญญัติหรือตัวแปรอันดับนั้น ๆ จะมีความถี่ของแต่ละประเภทหรืออันดับเท่าใด ดังนั้นระยะห่างสำหรับข้อมูลแบบผสมที่นำเสนอขึ้นใหม่ทั้งสองนี้ จึงเป็นอีกทางเลือกที่สามารถนำไปใช้ในการวิเคราะห์กลุ่มข้อมูลจริงได้อย่างมีประสิทธิภาพ

5.3 ข้อเสนอแนะ

การวิจัยนี้ ได้ทำการศึกษาภายใต้ขอบเขตของจำนวนตัวแปรอิสระ ค่าสัมประสิทธิ์สหสัมพันธ์ขนาดข้อมูลต่อกลุ่ม จำนวนกลุ่มข้อมูล จำนวนประเภทของตัวแปรนามบัญญัติ และจำนวนอันดับของตัวแปรอันดับ ซึ่งครอบคลุมเพียงบางกรณีเท่านั้น นอกจากนี้ยังเปรียบเทียบประสิทธิภาพระยะห่างจากการวิเคราะห์กลุ่มด้วยอัลกอริทึมจัดกลุ่มโดยรอบมีดอยด์เพียงอัลกอริทึมเดียวเท่านั้น ดังนั้นเพื่อประโยชน์ในการใช้งานจริง จึงควรมีการศึกษาเพิ่มเติม โดยขยายขอบเขตการวิจัยให้กว้างขึ้น เพื่อให้สามารถเลือกใช้ระยะห่างสำหรับวิเคราะห์กลุ่มที่เหมาะสมกับข้อมูลในชีวิตประจำวันได้ เพราะฉะนั้นผู้วิจัยจึงได้เสนอแนวทางในการศึกษาต่อ ดังนี้

1. ศึกษาจำนวนประเภทของตัวแปรนามบัญญัติ และจำนวนอันดับของตัวแปรอันดับ ที่หลากหลายขึ้น เพื่อให้สอดคล้องกับข้อมูลจริง เนื่องจากในทางปฏิบัติไม่สามารถควบคุมได้ว่าข้อมูลจะประกอบไปด้วยตัวแปรประเภทใด และมีจำนวนประเภทหรืออันดับเท่าใดบ้าง โดยข้อมูลชุดหนึ่งอาจมีทั้งตัวแปรนามบัญญัติที่มีจำนวนประเภทน้อย และตัวแปรอันดับที่มีจำนวนอันดับมากปะปนกัน
2. ศึกษากรณีที่จำนวนประเภทของตัวแปรนามบัญญัติ หรือจำนวนอันดับของตัวแปรอันดับ มีค่ามากกว่าจำนวนกลุ่มข้อมูล เพื่อหาระยะห่างที่ทำให้การวิเคราะห์กลุ่มข้อมูลลักษณะนี้มีประสิทธิภาพดี
3. ศึกษาอัลกอริทึมจัดกลุ่มอื่น ๆ ที่สามารถวิเคราะห์กลุ่มข้อมูลแบบผสมโดยประยุกต์ใช้กับระยะห่างที่ศึกษาได้ ซึ่งอาจทำให้การวิเคราะห์กลุ่มมีประสิทธิภาพดีกว่าการวิเคราะห์กลุ่มด้วยอัลกอริทึมจัดกลุ่มโดยรอบมีดอยด์ และเพื่อเป็นอีกตัวเลือกหนึ่งในการนำไปใช้วิเคราะห์กลุ่มข้อมูลแบบผสม

รายการอ้างอิง

Everitt, B. S., S. Landau, M. Leese and D. Stahl (2011). Cluster Analysis. London, A John Wiley and Sons, Ltd., Publication.

Gower, J. C. (1971). "A General Coefficient of Similarity and Some of Its Properties." Biometrics **27**(4): 857-871.

Kaufman, L. and P. J. Rousseeuw (1990). Finding Groups in Data: An Introduction to Cluster Analysis. USA, A Wiley-Interscience Publication.

Madhulatha, T. S. (2011). "Comparison between K-Means and K-Medoids Clustering Algorithms." Communications in Computer and Information Science **198**: 472-481.

Manning, C. D., P. Raghavan and H. Schutze (2009). An Introduction to Information Retrieval. Cambridge, Cambridge University Press.

Noorbehbahani, F., S. R. Mousavi and A. Mirzaei (2015). "An Incremental Mixed Data Clustering Method Using A New Distance Measure." Soft Computing **19**(3): 731-743.

Podani, J. (1999). "Extending Gower's General Coefficient of Similarity to Ordinal Characters." International Association for Plant Taxonomy **48**(2): 331-340.

Rand, W. M. (1971). "Objective Criteria for the Evaluation of Clustering Methods." Journal of the American Statistical Association **66**(336): 846-850.

Rousseeuw, P. J. (1987). "Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis." Jounal of Computational and Applied Mathematics **20**: 53-65.

ภาคผนวก

คำสั่งโปรแกรม R

การวิจัยครั้งนี้ ทำการจำลองข้อมูลและวิเคราะห์กลุ่มด้วยโปรแกรม R เวอร์ชัน 3.2.3 โดยเรียกใช้แพ็กเกจเพิ่มเติม ได้แก่ MASS cluster StatMatch clusteval และ Rcpp และดำเนินการเขียนชุดคำสั่งดังต่อไปนี้

ชุดคำสั่งสำหรับคำนวณระยะห่างของ P ระยะห่างแบบผสม KR&N และระยะห่างแบบผสม P&N

```
dfr <- function (a, b, fr){
        return((abs(fr[a] - fr[b]) + min(fr)) / max(fr[a], fr[b]))
       }

nbb.dist <- function(x){
        fr <- table(x)
        cdistv <- matrix(0, nrow=length(x), ncol=length(x))
        for (i in 1:(length(x)-1)){
                for (j in (i+1):length(x)){
                        if(is.na(x[i]) || is.na(x[j])){
                            cdistv[i,j] = 0
                        }else{
                                if(x[i] != x[j]){
                                        cdistv[i,j]= dfr(x[i], x[j], fr)
                                }
                        }
                cdistv[j,i] = cdistv[i,j]
                }
        }
        return(cdistv)
        }
```

```
## All P - KR&N - P&N Distances
## P        : Gower & Podani Distance
## KR&N  : Gower & KR & Nbb Distance
## P&N    : Gower & Podani & Nbb Distance

mix.dist <- function (data.x, method = c("P", "KR&N", "P&N"))
{
        mix.fcn <- function(x, method = c("P", "KR&N", "P&N")) {
                nx <- length(x)
                delta <- matrix(1, nx, nx)
                delta[outer(is.na(x), is.na(x), FUN = "|")] <- 0
                if (is.logical(x)) {
                        dd <- abs(outer(X = x, Y = x, FUN = "-"))
                }
                else if (is.character(x) || (is.factor(x) && !is.ordered(x))) {
                        if (method == "P") {
                                dd <- 1 - outer(X = x, Y = x, FUN = "==")
                        } else if (method == "KR&N" || method == "P&N"){
                                dd <- nbb.dist(x)
                        }
                }
                else if (is.ordered(x)) {
                        if (method == "P" || method == "P&N") {
                                x <- as.numeric(x)
                                x <- rank(x, ties.method = "average")
                                rng <- max(x) - min(x)
                                dd <- abs(outer(X = x, Y = x, FUN = "-"))/rng
                        } else if (method == "KR&N") {
                                x <- as.numeric(x)
                                rng <- max(x) - 1
                                zx <- (x - 1)/rng
```

```
                        dd <- abs(outer(X = zx, Y = zx, FUN = "-"))/(max(zx) - min(zx))
                }
        }
        else {
                rng <- max(x) - min(x)
                dd <- abs(outer(X = x, Y = x, FUN = "-"))/rng
        }
        list(dist = dd, delta = delta)
}

if (method != "P" && method != "KR&N" && method != "P&N")
        stop("there is no method.")
p <- ncol(data.x)
num <- array(0, c(nrow(data.x), nrow(data.x), p))
den <- array(0, c(nrow(data.x), nrow(data.x), p))
for (k in 1:p) {
        out.gow <- mix.fcn(x = data.x[, k], method)
        num[, , k] <- out.gow$dist * out.gow$delta
        den[, , k] <- out.gow$delta
}
out <- apply(num, c(1, 2), sum, na.rm = TRUE)/apply(den,
        c(1, 2), sum, na.rm = TRUE)
out
}
```

ชุดคำสั่งสำหรับจำลองข้อมูล

```
## Creating Mix Data for clustering
MixData <- function(N, Mu, Sigma, Cumulative, nom, support = list())
{
        cl <- length(N)
        p <- length(Mu[[1]])
        mixdata <- function (n, mu, sigma, cumulative, nom, support = list())
        {
                k <- length(cumulative)
                kj <- numeric(k)
                len <- length(support)
                p <- length(mu)
                for (i in 1:k) {
                        kj[i] <- length(cumulative[[i]]) + 1
                        if (len == 0) {
                                support[[i]] <- 1:kj[i]
                        }
                }
                mix <- mvrnorm(n, mu, sigma)
                if (n == 1)
                        mix <- matrix(mix, nrow = 1)
                for (i in 1:k) {
                        mix[, i] <- as.integer(cut(mix[, i], breaks = c(min(mix[,i]) - 1,
                                qnorm(cumulative[[i]],mu[i]), max(mix[, i]) + 1)))
                        mix[, i] <- support[[i]][mix[, i]]
                }
                mix <- data.frame(mix)
                if (nom == 0){
                        for (i in (nom+1):k){
                                mix[,i] <- ordered(mix[,i], levels = support[[i]])
```

```
                }
        }else if (nom == k){
                for (i in 1:nom){
                        mix[,i] <- factor(mix[,i])
                }
        }else{
                for (i in 1:nom){
                        mix[,i] <- factor(mix[,i])
                }
                for (i in (nom+1):k){
                        mix[,i] <- ordered(mix[,i], levels = support[[i]])
                }
        }
        return(mix)
}

data <- list()
for ( i in 1:cl ){
        mdata <- mixdata(N[i], Mu[[i]], Sigma[[i]], Cumulative[[i]], nom, support)
        data[[i]] <- mdata
}
data <- do.call("rbind", data)
data <- data.frame(data)
return(data)
}
```

ชุดคำสั่งสำหรับวิเคราะห์กลุ่มข้อมูล n ชุด

```
## Calculate purity
Purity <- function(clusters, classes) {
                sum(apply(table(clusters, classes), 2, max)) / length(clusters)
                }

## Generate data n rounds and calculate purity, rand, jaccard, silhouette, and time
MixCluster <- function(N, Mu, Sigma, Cumulative, nom, support, ord, k, cl, round)
{
        pur1 <- c(); pur2 <- c(); pur3 <- c(); pur4 <- c();
        rand1 <- c(); rand2 <- c(); rand3 <- c(); rand4 <- c();
        jac1 <- c(); jac2 <- c(); jac3 <- c(); jac4 <- c();
        silh1 <- c(); silh2 <- c(); silh3 <- c(); silh4 <- c();
        utime1 <- c(); utime2 <- c(); utime3 <- c(); utime4 <- c();
        stime1 <- c(); stime2 <- c(); stime3 <- c(); stime4 <- c();
        etime1 <- c(); etime2 <- c(); etime3 <- c(); etime4 <- c();

        if (nom != 0 && ord == 0){
            for (i in 1:round){
                Z <- MixData(N, Mu, Sigma, Cumulative, nom, support)
                time1 <- system.time(dist1 <- gower.dist(Z, KR.corr = T))
                time3 <- system.time(dist3 <- mix.dist(Z, method = "KR&N"))
                clust1 <- pam(dist1, k, diss=T)
                clust3 <- pam(dist3, k, diss=T)
                pur1 <- c(pur1, Purity(clust1$clustering,cl))
                pur3 <- c(pur3, Purity(clust3$clustering,cl))
                rand1 <- c(rand1, cluster_similarity(clust1$clustering,cl,similarity="rand"))
                rand3 <- c(rand3, cluster_similarity(clust3$clustering,cl,similarity="rand"))
                jac1 <- c(jac1, cluster_similarity(clust1$clustering,cl,similarity="jaccard"))
                jac3 <- c(jac3, cluster_similarity(clust3$clustering,cl,similarity="jaccard"))
```

```
        silh1 <- c(silh1, mean(silhouette(clust1)[,3]))
        silh3 <- c(silh3, mean(silhouette(clust3)[,3]))
        utime1 <- c(utime1, time1[[1]])
        utime3 <- c(utime3, time3[[1]])
        stime1 <- c(stime1, time1[[2]])
        stime3 <- c(stime3, time3[[2]])
        etime1 <- c(etime1, time1[[3]])
        etime3 <- c(etime3, time3[[3]])
    }
    output <- cbind(pur1, pur3, rand1, rand3, jac1, jac3, silh1, silh3, utime1,
        utime3, stime1, stime3, etime1, etime3)

}else if (nom == 0 && ord != 0){
    for (i in 1:round){
        Z <- MixData(N, Mu, Sigma, Cumulative, nom, support)
        time1 <- system.time(dist1 <- gower.dist(Z, KR.corr = T))
        time2 <- system.time(dist2 <- mix.dist(Z, method = "P"))
        clust1 <- pam(dist1, k, diss=T)
        clust2 <- pam(dist2, k, diss=T)
        pur1 <- c(pur1,Purity(clust1$clustering,cl))
        pur2 <- c(pur2,Purity(clust2$clustering,cl))
        rand1 <- c(rand1,cluster_similarity(clust1$clustering,cl,similarity="rand"))
        rand2 <- c(rand2,cluster_similarity(clust2$clustering,cl,similarity="rand"))
        jac1 <- c(jac1,cluster_similarity(clust1$clustering,cl,similarity="jaccard"))
        jac2 <- c(jac2,cluster_similarity(clust2$clustering,cl,similarity="jaccard"))
        silh1 <- c(silh1, mean(silhouette(clust1)[,3]))
        silh2 <- c(silh2, mean(silhouette(clust2)[,3]))
        utime1 <- c(utime1, time1[[1]])
        utime2 <- c(utime2, time2[[1]])
        stime1 <- c(stime1, time1[[2]])
        stime2 <- c(stime2, time2[[2]])
```

```
        etime1 <- c(etime1, time1[[3]])
        etime2 <- c(etime2, time2[[3]])
    }
    output <- cbind(pur1, pur2, rand1, rand2, jac1, jac2, silh1, silh2,utime1,
        utime2, stime1, stime2,etime1, etime2)

}else{
    for (i in 1:round){
        Z <- MixData(N, Mu, Sigma, Cumulative, nom, support)
        time1 <- system.time(dist1 <- gower.dist(Z, KR.corr = T))
        time2 <- system.time(dist2 <- mix.dist(Z, method = "P"))
        time3 <- system.time(dist3 <- mix.dist(Z, method = "KR&N"))
        time4 <- system.time(dist4 <- mix.dist(Z, method = "P&N"))
        clust1 <- pam(dist1, k, diss=T)
        clust2 <- pam(dist2, k, diss=T)
        clust3 <- pam(dist3, k, diss=T)
        clust4 <- pam(dist4, k, diss=T)
        pur1 <- c(pur1,Purity(clust1$clustering,cl))
        pur2 <- c(pur2,Purity(clust2$clustering,cl))
        pur3 <- c(pur3,Purity(clust3$clustering,cl))
        pur4 <- c(pur4,Purity(clust4$clustering,cl))
        rand1 <- c(rand1,cluster_similarity(clust1$clustering,cl,similarity="rand"))
        rand2 <- c(rand2,cluster_similarity(clust2$clustering,cl,similarity="rand"))
        rand3 <- c(rand3,cluster_similarity(clust3$clustering,cl,similarity="rand"))
        rand4 <- c(rand4,cluster_similarity(clust4$clustering,cl,similarity="rand"))
        jac1 <- c(jac1,cluster_similarity(clust1$clustering,cl,similarity="jaccard"))
        jac2 <- c(jac2,cluster_similarity(clust2$clustering,cl,similarity="jaccard"))
        jac3 <- c(jac3,cluster_similarity(clust3$clustering,cl,similarity="jaccard"))
        jac4 <- c(jac4,cluster_similarity(clust4$clustering,cl,similarity="jaccard"))
        silh1 <- c(silh1, mean(silhouette(clust1)[,3]))
        silh2 <- c(silh2, mean(silhouette(clust2)[,3]))
```

```
            silh3 <- c(silh3, mean(silhouette(clust3)[,3]))
            silh4 <- c(silh4, mean(silhouette(clust4)[,3]))
            utime1 <- c(utime1, time1[[1]])
            utime2 <- c(utime2, time2[[1]])
            utime3 <- c(utime3, time3[[1]])
            utime4 <- c(utime4, time4[[1]])
            stime1 <- c(stime1, time1[[2]])
            stime2 <- c(stime2, time2[[2]])
            stime3 <- c(stime3, time3[[2]])
            stime4 <- c(stime4, time4[[2]])
            etime1 <- c(etime1, time1[[3]])
            etime2 <- c(etime2, time2[[3]])
            etime3 <- c(etime3, time3[[3]])
            etime4 <- c(etime4, time4[[3]])
        }
        output <-  cbind(pur1, pur2, pur3, pur4, rand1, rand2, rand3, rand4,
                   jac1, jac2, jac3, jac4, silh1, silh2, silh3, silh4,
                   utime1, utime2, utime3, utime4, stime1, stime2, stime3, stime4,
                   etime1, etime2, etime3, etime4)
    }
    return(output)
}
```

ชุดคำสั่งสำหรับกำหนดค่าพารามิเตอร์เริ่มต้น

กรณีตัวอย่าง ข้อมูลรูปแบบที่ 1 เมื่อ K = 3, ρ = 0.2 และ n = 20

```
round = 1000
nom = 3
ord = 3
R2 <- matrix(c(1.0,0.2,0.2,0.2,0.2,0.2,0.2,0.2,0.2,
               0.2,1.0,0.2,0.2,0.2,0.2,0.2,0.2,0.2,
               0.2,0.2,1.0,0.2,0.2,0.2,0.2,0.2,0.2,
               0.2,0.2,0.2,1.0,0.2,0.2,0.2,0.2,0.2,
               0.2,0.2,0.2,0.2,1.0,0.2,0.2,0.2,0.2,
               0.2,0.2,0.2,0.2,0.2,1.0,0.2,0.2,0.2,
               0.2,0.2,0.2,0.2,0.2,0.2,1.0,0.2,0.2,
               0.2,0.2,0.2,0.2,0.2,0.2,0.2,1.0,0.2,
               0.2,0.2,0.2,0.2,0.2,0.2,0.2,0.2,1.0),9)
R8 <- matrix(c(1.0,0.8,0.8,0.8,0.8,0.8,0.8,0.8,0.8,
               0.8,1.0,0.8,0.8,0.8,0.8,0.8,0.8,0.8,
               0.8,0.8,1.0,0.8,0.8,0.8,0.8,0.8,0.8,
               0.8,0.8,0.8,1.0,0.8,0.8,0.8,0.8,0.8,
               0.8,0.8,0.8,0.8,1.0,0.8,0.8,0.8,0.8,
               0.8,0.8,0.8,0.8,0.8,1.0,0.8,0.8,0.8,
               0.8,0.8,0.8,0.8,0.8,0.8,1.0,0.8,0.8,
               0.8,0.8,0.8,0.8,0.8,0.8,0.8,1.0,0.8,
               0.8,0.8,0.8,0.8,0.8,0.8,0.8,0.8,1.0),9)
V <- matrix(c(1,0,0,0,0,0,0,0,0,
              0,1,0,0,0,0,0,0,0,
              0,0,1,0,0,0,0,0,0,
              0,0,0,1,0,0,0,0,0,
              0,0,0,0,1,0,0,0,0,
              0,0,0,0,0,1,0,0,0,
              0,0,0,0,0,0,1,0,0,
```

```
                    0,0,0,0,0,0,0,1,0,
                    0,0,0,0,0,0,0,0,1),9)
## Label
lb1 <- c(1,2,3,4,5)
lb2 <- c(1,2,4,6,8)
#### case A :
Label <- list(lb1, lb1, lb1, lb1, lb1, lb1)
####
####
###### case A.1 : 9 vars
###### Mean
###### for 3 nom & 3 ord & 3 num
mu1 <- c(0,0,0,0,0,0,1,2,3)
mu2 <- c(0,0,0,0,0,0,4,5,6)
mu3 <- c(0,0,0,0,0,0,7,8,9)
mu4 <- c(0,0,0,0,0,0,10,11,12)
mu5 <- c(0,0,0,0,0,0,13,14,15)
###### Probability
###### vars w/ 5 levels
cu51 <- c(0.70, 0.80, 0.90, 0.95)
cu52 <- c(0.05, 0.75, 0.85, 0.95)
cu53 <- c(0.05, 0.10, 0.80, 0.90)
cu54 <- c(0.10, 0.15, 0.20, 0.90)
cu55 <- c(0.10, 0.20, 0.25, 0.30)
###### for 3 nom & 3 ord
cumulative1 <- list(cu51,cu51,cu51,cu51,cu51,cu51)
cumulative2 <- list(cu52,cu52,cu52,cu52,cu52,cu52)
cumulative3 <- list(cu53,cu53,cu53,cu53,cu53,cu53)
cumulative4 <- list(cu54,cu54,cu54,cu54,cu54,cu54)
cumulative5 <- list(cu55,cu55,cu55,cu55,cu55,cu55)
```

```
########  case A.1.1 : k = 3
k <- 3
Mu <- list(mu1, mu2, mu3)
Cumulative <- list(cumulative1, cumulative2, cumulative3)
########## case A.1.1.1 : rho = 0.2
VC <- V%*%R2%*%V
Sigma <- list(VC,VC,VC)
########### case A.1.1.1.1 : n = 20
N <- c(20, 20, 20)
cl <- c(rep(1,20), rep(2,20), rep(3,20))
ResultA.1.1.1.1 <- MixCluster(N, Mu, Sigma, Cumulative, nom, support = Label, ord, k,
                              cl, round)
```

ประวัติผู้เขียนวิทยานิพนธ์

นางสาวพิชญา บุตรขุนทอง เกิดวันที่ 15 พฤษภาคม พ.ศ.2533 สำเร็จการศึกษาปริญญาวิทยาศาสตร์บัณฑิต (วท.บ.) สาขาวิชาคณิตศาสตร์ ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2555 และเข้าศึกษาต่อในหลักสูตรวิทยาศาสตร์มหาบัณฑิต (วท.ม.) สาขาวิชาสถิติ ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2556