บทที่ 5

สรุปผลการวิจัยและข้อเสนอแนะ

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาและเปรียบเทียบประสิทธิภาพในการจำแนกกลุ่มแบบอันดับ (Ordered-group) ของข้อมูลระหว่างวิธีการวิเคราะห์จำแนกประเภท (Discriminant Analysis: DA) และวิธีการวิเคราะห์ความถดถอยโลจิสติกอันดับ (Ordinal Logistic Regression: OLR) เมื่อข้อมูลตัวแปรอิสระมีการแจกแจงแบบปกติหลายตัวแปร (Multivariate Normal) โดยมีเงื่อนไขที่ต้องการศึกษาและเปรียบเทียบดังนี้

(1) ขนาดตัวอย่าง 120 , 240 และ 360

(2) ขนาดตัวอย่างในแต่ละกลุ่มถูกแบ่งศึกษาออกเป็น กรณีสมดุล , สมดุลปานกลาง และไม่สมดุล

(3) จำนวนกลุ่มอันดับที่จะจำแนกเป็น 3 , 4 และ 5 กลุ่ม

(4) กรณีตัวแปรอิสระไม่สัมพันธ์กันและสัมพันธ์กันด้วยค่าสหสัมพันธ์ (Correlation : ρ) เป็น 0.1, 0.5 และ 0.9

(5) กรณีเมทริกซ์ค่าความแปรปรวนร่วมของตัวแปรอิสระของแต่ละกลุ่มเท่ากันและไม่เท่ากัน

เกณฑ์ที่ใช้ในการตัดสินใจและเปรียบเทียบคือ อัตราความผิดพลาดที่เห็นชัด (Apparent Rate Error: APER) ซึ่งคำนวณตามจำนวนตัวอย่างที่ถูกจำแนกผิดกลุ่มในหน่วยเปอร์เซ็นต์

การศึกษาวิจัยใช้ข้อมูลที่ได้จากการจำลองด้วยเทคนิคมอนติคาร์โล สำหรับสถานการณ์ต่างๆ ที่กำหนดโดยทำการทดลองซ้ำ 1000 ครั้งในแต่ละสถานการณ์วิจัย

5.1 สรุปผลการวิจัย

จากผลการวิจัยได้ข้อสรุปดังนี้

1. เมื่อศึกษาปัจจัยทั้ง 5 ได้แก่ จำนวนตัวแปรอิสระ, ขนาดตัวอย่าง, จำนวนตัวอย่างในแต่ละกลุ่ม, จำนวนกลุ่ม และความสัมพันธ์ระหว่างตัวแปรอิสระ พบว่าทั้ง 5 ปัจจัยมีผลต่อการจำแนกกลุ่มแบบอันดับของทั้งวิธีวิเคราะห์จำแนกประเภทและวิธีวิเคราะห์ความถดถอยโลจิสติกอันดับ ดังที่ได้แสดงค่าสหสัมพันธ์เชิงส่วน (Partial Correlation) ในตารางที่ 4.24 – 4.27 สรุปแยกเป็นกรณีดังตารางที่ 5.1

ตารางที่ 5.1 ตารางสรุปความสำคัญของปัจจัยที่มีผลต่อการจำแนกกลุ่มของการวิเคราะห์ทั้งสองวิธี

ระดับความสำคัญ	กรณีเมทริกซ์ความแปรปรวนร่วมเท่ากัน		กรณีเมทริกซ์ความแปรปรวนร่วมไม่เท่ากัน	
	วิธี DA	วิธี OLR	วิธี DA	วิธี OLR
ความสำคัญมากสุด	จำนวนกลุ่ม	จำนวนกลุ่ม	จำนวนกลุ่ม	จำนวนกลุ่ม
↓	ขนาดตัวอย่าง	ขนาดตัวอย่าง	ความสัมพันธ์ระหว่างตัวแปรอิสระ	จำนวนตัวแปรอิสระ
↓	ความสัมพันธ์ระหว่างตัวแปรอิสระ	ความสัมพันธ์ระหว่างตัวแปรอิสระ	จำนวนตัวแปรอิสระ	ขนาดตัวอย่าง
↓	จำนวนตัวแปรอิสระ	จำนวนตัวแปรอิสระ	ขนาดตัวอย่าง	ความสัมพันธ์ระหว่างตัวแปรอิสระ
ความสำคัญน้อยสุด	จำนวนตัวอย่างในแต่ละกลุ่ม	จำนวนตัวอย่างในแต่ละกลุ่ม	จำนวนตัวอย่างในแต่ละกลุ่ม	จำนวนตัวอย่างในแต่ละกลุ่ม

DA: การวิเคราะห์จำแนกประเภท (Discriminant Analysis)

OLR: การวิเคราะห์ความถดถอยโลจิสติกอันดับ (Ordinal Logistic Regression)

ทั้งนี้จะพบว่าจำนวนกลุ่มเป็นปัจจัยที่สำคัญสุดต่อความผิดพลาดในการจำแนกกลุ่มทั้งนี้อาจจะเป็นเพราะว่าจำนวนกลุ่มที่เพิ่มขึ้นจะทำให้วิธีวิเคราะห์ทั้งสองวิธีต้องสร้างสมการจำแนกกลุ่มที่เพิ่มขึ้นนั่นทำให้ค่าความผิดพลาด (error) ต่างๆจากการประมาณค่าพารามิเตอร์มากขึ้นตามไปด้วย ในขณะที่จำนวนตัวอย่างในแต่ละกลุ่มไม่ส่งผลมากนักเนื่องจากงานวิจัยนี้กำหนดจำนวนตัวอย่างค่อนข้างไม่แตกต่างกันมากนักแม้ว่าจะเป็นกรณีไม่สมดุลก็ตามดังนั้นจำนวนตัวอย่างที่แตกต่างกันจึงอาจไม่ส่งผลให้เห็นชัดเจน

2. เมื่อเปรียบเทียบปัจจัยทั้ง 5 ได้แก่ จำนวนตัวแปรอิสระ, ขนาดตัวอย่าง, จำนวนตัวอย่างในแต่ละกลุ่ม, จำนวนกลุ่ม และความสัมพันธ์ระหว่างตัวแปรอิสระ แต่ละปัจจัยพบว่า ทั้งวิธีวิเคราะห์จำแนกประเภทและวิธีวิเคราะห์ความถดถอยโลจิสติกอันดับให้แนวโน้มไปในทางเดียวกัน ดังนี้

<u>ปัจจัยด้านจำนวนกลุ่ม</u>

เมื่อจำนวนกลุ่มมากขึ้นทั้งสองวิธีจะจำแนกกลุ่มผิดพลาดมากขึ้น ทั้งนี้เนื่องจากเมื่อจำนวนกลุ่มมากขึ้นจะมีสมการในการจำแนกมากขึ้นทั้งสองวิธีซึ่งส่งผลให้ค่าความผิดพลาด (error) ต่างๆ ในการประมาณค่าพารามิเตอร์ต่างๆมากขึ้นตามไปด้วย

<u>ปัจจัยด้านขนาดตัวอย่าง</u>

เมื่อขนาดตัวอย่างมากขึ้นทั้งสองวิธีจะจำแนกกลุ่มได้ผิดพลาดลดลง เป็นผลที่สอดคล้องกับหลักทางสถิติทั่วไป ซึ่งขนาดตัวอย่างที่มากขึ้นจะช่วยลดความแปรปรวนของข้อมูลและการวิเคราะห์ลง

<u>ปัจจัยด้านความสัมพันธ์ของตัวแปรอิสระ</u>

เมื่อตัวแปรอิสระมีความสัมพันธ์กันทั้งสองวิธีจะจำแนกกลุ่มผิดพลาดมากขึ้นกล่าวคือที่สหสัมพันธ์ของตัวแปรอิสระเท่ากับ 0.9 ทั้งสองวิธีจะจำแนกกลุ่มผิดพลาดมากที่สุด รองลงมาคือที่สหสัมพันธ์ 0.5, 0.1 และเมื่อตัวแปรอิสระไม่สัมพันธ์กันทั้งสองวิธีให้ค่าผิดพลาดของการจำแนกต่ำที่สุด ปัญหาด้านความสัมพันธ์ของตัวแปรอิสระนั้นเป็นข้อห้ามสำคัญสำหรับการวิเคราะห์ความถดถอย (regression) อยู่แล้ว โดยเฉพาะการวิเคราะห์โลจิสติกที่ถือเป็นข้อสมมติเบื้องต้น

<u>ปัจจัยด้านจำนวนตัวแปรอิสระ</u>

เมื่อจำนวนตัวแปรอิสระมากขึ้นทั้งสองวิธีจะจำแนกกลุ่มผิดพลาดมากขึ้น เนื่องจากตัวแปรอิสระที่มากขึ้นนั้นย่อมทำให้เกิดค่าคลาดเคลื่อนหรือความผิดพลาด (error) ต่างๆในการประมาณพารามิเตอร์ที่มากตามไปด้วย

<u>ปัจจัยด้านจำนวนตัวอย่างในแต่ละกลุ่ม</u>

ที่จำนวนตัวอย่างแบบไม่สมดุลทั้งสองวิธีจะให้การจำแนกกลุ่มที่ผิดพลาดมากกว่าที่ระดับสมดุลปานกลางและที่ระดับสมดุล แต่การเพิ่มขึ้นของความผิดพลาดในการจำแนกกลุ่มนั้นเพิ่มขึ้นเพียงเล็กน้อยเท่านั้นหรือกล่าวได้ว่าส่งผลเพียงเล็กน้อยเท่านั้นโดยดูได้จากค่าสหสัมพันธ์เชิงส่วนที่แสดงให้เห็นว่า ปัจจัยด้านจำนวนตัวอย่างในแต่ละกลุ่ม มีผลต่อความผิดพลาดในการจำแนกน้อยที่สุดโดยทั้งนี้อาจเป็นเพราะว่าแม้ว่าจะเป็นระดับไม่สมดุลแต่ขนาดตัวอย่างที่กำหนดในแต่ละกลุ่มก็ไม่แตกต่างกันมากนัก อาจจะยังอยู่ในระดับที่ไม่ส่งผลต่อการจำแนก

3. เมื่อศึกษากรณีการไม่เท่ากันของเมทริกซ์ความแปรปรวนร่วมของตัวแปรอิสระของทุกกลุ่มซึ่งเป็นการศึกษาเมื่อข้อตกลงเบื้องต้นของการวิเคราะห์จำแนกประเภทไม่เป็นจริงจะพบว่าวิธีวิเคราะห์จำแนกประเภทจะจำแนกกลุ่มผิดพลาดมากขึ้นในทุกกรณีโดยเฉลี่ยแล้ว 25.53% ในขณะที่วิธีวิเคราะห์ความถดถอยโลจิสติกอันดับซึ่งไม่มีสมมติฐานนี้เป็นข้อกำหนดจะจำแนกกลุ่มผิดพลาดมากขึ้นเฉลี่ยแล้ว 5.60%

4. จากการศึกษาเปรียบเทียบในทุกปัจจัยและทุกเงื่อนไขของการวิจัยจะพบว่าวิธีวิเคราะห์จำแนกประเภทจะจำแนกกลุ่มได้ผิดพลาดมากกว่าวิธีวิเคราะห์ความถดถอยโลจิสติกอันดับ ทั้งนี้อาจเป็นเพราะว่าวิธีวิเคราะห์ความถดถอยโลจิสติกอันดับมีลักษณะตัวแบบ (model) ที่วิเคราะห์โดยการเปรียบเทียบความน่าจะเป็นทีละอันดับจึงให้ความแม่นยำในการจำแนกมากกว่าวิธีวิเคราะห์จำแนกประเภทซึ่งรวมข้อมูลเข้าด้วยการในการสร้างตัวแบบ

ดังนั้นจึงสรุปได้ว่าวิธีวิเคราะห์ความถดถอยโลจิสติกอันดับ (Ordinal Logistic Regression) จะเป็นวิธีที่มีประสิทธิภาพในการจำแนกกลุ่มแบบอันดับดีกว่าวิธีวิเคราะห์จำแนกประเภท (Discriminant Analysis) ภายใต้เงื่อนไขของการวิจัยครั้งนี้

การตัดสินใจเลือกตัวแบบสำหรับการวิเคราะห์จำแนกกลุ่มแบบอันดับนั้น จากผลการวิจัยจึงเห็นสมควรสรุปได้ว่าเราควรจะเลือกใช้วิธีวิเคราะห์ความถดถอยโลจิสติกมากกว่าวิธีวิเคราะห์จำแนกประเภท ทั้งนี้การวิจัยนี้ทำภายใต้ข้อสมมติการแจกแจงของตัวแปรอิสระเป็นแบบปกติหลายตัวแปรอันเป็นข้อกำหนดเบื้องต้นของวิธีวิเคราะห์จำแนกประเภท ซึ่งในความเป็นจริงข้อมูลของเราอาจจะไม่ได้มีการแจกแจงอย่างที่สมมติขึ้น ซึ่งในกรณีนั้นก็จะยิ่งเหมาะกับวิธีการวิเคราะห์ความถดถอยโลจิสติกอันดับซึ่งถูกแสดงให้เห็นแล้วว่าสามารถจำแนกกลุ่มได้ดีกว่าวิธีวิเคราะห์จำแนกประเภทครอบคลุมหลายรูปแบบของข้อมูล

5.2 ข้อเสนอแนะ

การวิจัยครั้งนี้ศึกษาภายใต้เงื่อนไขที่ตัวแปรอิสระมีการแจกแจงแบบปกติหลายตัวแปรซึ่งเป็นข้อสมมติพื้นฐานสำหรับการวิเคราะห์จำแนกประเภท โดยในงานวิจัยต่อไปอาจจะศึกษาข้อมูลตัวแปรอิสระที่มีการแจกแจงในรูปแบบอื่นโดยอาจอ้างอิงจากข้อมูลจริงเฉพาะด้านใดด้านหนึ่งเช่น ข้อมูลด้านการแพทย์ ข้อมูลด้านเศรษฐกิจเป็นต้น การศึกษาในกรณีขนาดตัวอย่างน้อยหรือข้อมูลหายาก รวมถึงการเลือกใช้ตัวแบบอื่นสำหรับการวิเคราะห์ความถดถอยโลจิสติกอันดับ เช่น baseline logit model , adjacent-category logistic model หรือ continuation-ratio logistic model อีกทั้งอาจจะศึกษาเงื่อนไขอื่นที่เหมาะสมสำหรับการวิเคราะห์จำแนกกลุ่มแบบอันดับเพื่อลดค่าผิดพลาดในการจำแนกต่อไป