

## บทที่ 5

### สรุปผลการวิจัย

จากการเปรียบเทียบประสิทธิภาพการทำนายการเป็นสมาชิกกลุ่มที่ถูกตัดด้วยวิธีวิเคราะห์การถดถอยพหุคูณคุณ วิธีวิธีดัจรีเกรสชัน วิธีวิเคราะห์การถดถอยโลจิสติก และวิธีวิเคราะห์จำแนกกลุ่มในเชิงเส้นตรง โดยใช้ค่าเฉลี่ยของความคลาดเคลื่อนในการจำแนกกลุ่ม ค่าเฉลี่ยของ B และค่าเฉลี่ยของ C ซึ่งสามารถสรุปได้ดังนี้

#### 5.1 การเปรียบเทียบประสิทธิภาพการจำแนกกลุ่มโดยใช้ค่าเฉลี่ยของความคลาดเคลื่อนในการจำแนกกลุ่ม

กรณีมีตัวแปรที่ใช้ทำนาย 4 ตัว ทุกวิธีให้ค่าเฉลี่ยของความคลาดเคลื่อนในการจำแนกกลุ่มไม่แตกต่างกัน

##### กรณีมีตัวแปรที่ใช้ทำนาย 6 ตัว

1) กรณีไม่มีตัวแปรเชิงคุณภาพและกรณีมีตัวแปรเชิงคุณภาพ 2 ตัว ทุกวิธีให้ค่าเฉลี่ยของความคลาดเคลื่อนในการจำแนกกลุ่มไม่แตกต่างกัน ในทุกระดับของ VIF

2) กรณีมีตัวแปรเชิงคุณภาพ 3 ตัว และ  $VIF \leq 4$  วิธีวิเคราะห์การถดถอยโลจิสติกให้ค่าเฉลี่ยของความคลาดเคลื่อนในการจำแนกกลุ่มน้อยกว่าวิธี Ridge 1 และวิธี Ridge 2 ในกรณี  $4 < VIF \leq 10$  วิธีวิเคราะห์การถดถอยโลจิสติกให้ค่าเฉลี่ยของความคลาดเคลื่อนในการจำแนกกลุ่มน้อยกว่าวิธี Ridge1 ในกรณี  $VIF > 10$  ทุกวิธีให้ค่าเฉลี่ยของความคลาดเคลื่อนในการจำแนกกลุ่มไม่แตกต่างกัน

3) กรณีมีตัวแปรเชิงคุณภาพ 4 ตัว วิธีวิเคราะห์การถดถอยโลจิสติกให้ค่าเฉลี่ยของความคลาดเคลื่อนในการจำแนกกลุ่มน้อยกว่าวิธีวิเคราะห์จำแนกกลุ่มในเชิงเส้นตรง วิธีวิเคราะห์การถดถอยพหุคูณและวิธี Ridge 1 ในทุกระดับของ VIF

##### กรณีมีตัวแปรที่ใช้ทำนาย 10 ตัว

1) กรณีไม่มีตัวแปรเชิงคุณภาพและกรณีมีตัวแปรเชิงคุณภาพ 2 ตัว ทุกวิธีให้ค่าเฉลี่ยของความคลาดเคลื่อนในการจำแนกกลุ่มไม่แตกต่างกัน ในทุกระดับของ VIF

2) กรณีมีตัวแปรเชิงคุณภาพ 4 ตัว 5 ตัว และ 6 ตัว วิธีวิเคราะห์การถดถอยโลจิสติกให้ค่าเฉลี่ยของความคลาดเคลื่อนในการจำแนกกลุ่มน้อยกว่าวิธีวิเคราะห์จำแนกกลุ่มในเชิงเส้นตรง วิธีวิเคราะห์การถดถอยพหุคูณและวิธีวิธีดัจรีเกรสชัน (Ridge1,Ridge2)

จากการศึกษาพบว่าในกรณีที่ VIF ของตัวแปรที่ใช้ทำนายซึ่งมีค่ามากที่สุดมีค่ามากขึ้น ทุกวิธีให้ค่าเฉลี่ยของความคลาดเคลื่อนในการจำแนกมีค่าลดลง แต่ในกรณีที่ตัวแปรที่ใช้ทำนาย 10 ตัวโดยเป็นตัวแปรเชิงคุณภาพ 6 ตัว วิธีวิธีรีเกรสชันให้ค่าเฉลี่ยของความคลาดเคลื่อนในการจำแนกไม่แตกต่างกัน

จากการเปรียบเทียบประสิทธิภาพในการทำนายการเป็นสมาชิกกลุ่มจากการใช้วิธีวิเคราะห์การถดถอยโลจิสติก วิธีวิเคราะห์จำแนกกลุ่มในเชิงเส้นตรง วิธีวิเคราะห์การถดถอยเชิงพหุและวิธีวิธีรีเกรสชัน โดยพิจารณาจากค่าเฉลี่ยของความคลาดเคลื่อนในการจำแนกกลุ่มซึ่งพบว่าวิธีวิเคราะห์การถดถอยโลจิสติกให้ค่าเฉลี่ยของความคลาดเคลื่อนน้อยที่สุด แต่จากการทดสอบสมมติฐานความแตกต่างระหว่างค่าเฉลี่ยของความคลาดเคลื่อนที่ได้จากวิธีวิเคราะห์การถดถอยโลจิสติกกับวิธีวิเคราะห์จำแนกกลุ่มในเชิงเส้นตรง วิธีวิธีรีเกรสชัน และวิธีวิเคราะห์การถดถอยพหุคูณ โดยใช้ระดับนัยสำคัญ 0.05 พบว่าในทุกกรณีวิธีวิเคราะห์การถดถอยโลจิสติกให้ค่าเฉลี่ยของความคลาดเคลื่อนในการจำแนกไม่แตกต่างจากวิธีวิเคราะห์จำแนกกลุ่มในเชิงเส้นตรง ซึ่งอธิบายได้ว่ามีความสามารถในการจำแนกกลุ่มไม่แตกต่างกัน แต่ในบางกรณีวิธีวิเคราะห์การถดถอยโลจิสติกก็มีความสามารถในการจำแนกไม่แตกต่างจากวิธีวิธีรีเกรสชันและวิธีวิเคราะห์การถดถอยพหุคูณ ซึ่งสามารถสรุปผลการศึกษาได้ดังนี้

ตาราง 5.1 วิธีที่ให้ค่าเฉลี่ยของความคลาดเคลื่อนในการจำแนกกลุ่มน้อยที่สุดจำแนกตามจำนวนตัวแปรที่ใช้ทำนายและจำนวนตัวแปรเชิงคุณภาพและระดับ VIF

| จำนวนตัวแปรที่ใช้ทำนาย | จำนวนตัวแปรเชิงคุณภาพ | VIF          | วิธีที่ควรเลือกใช้ |     |     |        |        |
|------------------------|-----------------------|--------------|--------------------|-----|-----|--------|--------|
|                        |                       |              | Logit              | LDA | OLS | Ridge1 | Ridge2 |
| 4                      | 0 , 1 , 2             | ทุกระดับ     | ✓                  | ✓   | ✓   | ✓      | ✓      |
| 6                      | 0 , 2                 | ทุกระดับ     | ✓                  | ✓   | ✓   | ✓      | ✓      |
|                        |                       | VIF ≤ 4      | ✓                  | ✓   | ✓   |        |        |
|                        | 3                     | 4 < VIF ≤ 10 | ✓                  | ✓   | ✓   |        |        |
|                        |                       | VIF > 10     | ✓                  | ✓   | ✓   | ✓      | ✓      |
|                        | 4                     | ทุกระดับ     | ✓                  | ✓   |     |        |        |
| 10                     | 0 , 2                 | ทุกระดับ     | ✓                  | ✓   |     |        |        |
|                        | 4 , 5 , 6             | ทุกระดับ     | ✓                  | ✓   |     |        |        |

## 5.2 การเปรียบเทียบประสิทธิภาพการจำแนกกลุ่มโดยใช้ ค่าเฉลี่ยของ B

ค่า B คำนวณจากสูตร  $B = 1 - \sum_{i=1}^n (P_i - Y_i)^2 / n$  โดย  $P_i$  เป็นค่า

ความน่าจะเป็นในการจำแนกตัวแปรตามให้อยู่ในกลุ่ม  $i$   $Y_i$  เป็นค่าของตัวแปรตามซึ่งมีค่าเท่ากับ 0 หรือ 1  $n$  เป็นจำนวนสมาชิกของตัวแปร  $Y_i$   $B$  มีค่าอยู่ในช่วง  $[0, 1]$  ซึ่งหากมีค่าเท่ากับ 1 อธิบายได้ว่าสามารถอธิบายการจำแนกได้ถูกต้องสมบูรณ์ ในกรณีที่  $n_0 = n_1$   $B$  มีค่าเท่ากับ 0.75 ค่า  $B$  นอกจากจะใช้วัดความสามารถในการจำแนกแล้วยังสามารถวัดความแม่นยำในการพยากรณ์ (accuracy of prediction)

จากผลการศึกษาสามารถสรุปได้ว่าวิธีวิเคราะห์การถดถอยโลจิสติกให้ค่าเฉลี่ยของ  $B$  ไม่แตกต่างจากวิธีวิเคราะห์จำแนกกลุ่มในเชิงเส้นตรง นอกจากนี้ยังพบว่าสองวิธีนี้ให้ค่าเฉลี่ยของ  $B$  มากที่สุดในทุกกรณี เมื่อพิจารณาในแต่ละระดับของจำนวนตัวแปรที่ใช้ทำนายและจำนวนตัวแปรเชิงคุณภาพพบว่าเมื่อข้อมูลมีค่า VIF เพิ่มขึ้นจะทำให้ค่าเฉลี่ยของ  $B$  เพิ่มขึ้น เมื่อจำนวนตัวแปรที่ใช้ทำนายเพิ่มขึ้นจะทำให้ค่าเฉลี่ยของ  $B$  เพิ่มขึ้น เมื่อพิจารณาที่แต่ละระดับของจำนวนตัวแปรที่ใช้ทำนายและค่า VIF พบว่าในกรณีที่ประชากรมีจำนวนตัวแปรเชิงคุณภาพเพิ่มขึ้นจะทำให้ค่าเฉลี่ยของ  $B$  เพิ่มขึ้น

## 5.3 การเปรียบเทียบประสิทธิภาพการจำแนกกลุ่มโดยใช้ ค่าเฉลี่ยของ C

ค่า C คำนวณจากสูตร  $C = \sum_{i=1}^{n_0} \sum_{j=1}^{n_1} \left[ I(P_j > P_i) + \frac{1}{2} I(P_j = P_i) \right] / n_0 n_1$

เป็นค่าวัดความสามารถในการจำแนกค่าของตัวแปรตาม โดย  $n_0$  เป็นจำนวนตัวแปรสุ่ม  $Y_i$  ที่มีค่าเท่ากับ 0  $n_1$  เป็นจำนวนตัวแปรสุ่ม  $Y_i$  ที่มีค่าเท่ากับ 1  $P_k$  เป็นค่าประมาณของ  $P(Y_k=1|X_k)$  จาก  $P(Y_i=1|X_i) = \frac{e^{\beta^T X_i}}{1 + e^{\beta^T X_i}}$  โดย  $Y_i$  เป็นตัวแปรสุ่มที่เป็นอิสระกันและมีการแจกแจงเบอร์นูลลี  $\beta$  เป็นค่าประมาณสัมประสิทธิ์การถดถอย ซึ่งหาก  $C$  มีค่าเท่ากับ 1 อธิบายได้ว่าสามารถอธิบายการจำแนกได้ถูกต้องสมบูรณ์

จากผลการศึกษาสามารถสรุปได้ว่าวิธีวิเคราะห์การถดถอยโลจิสติกให้ค่าเฉลี่ยของ  $C$  ไม่แตกต่างจากวิธีวิเคราะห์จำแนกกลุ่มในเชิงเส้นตรง วิธีวิธีจี้เรชันและวิธีวิเคราะห์การถดถอยพหุคูณในทุกกรณี เมื่อพิจารณาในแต่ละระดับของจำนวนตัวแปรที่ใช้ทำนายและจำนวนตัวแปรเชิงคุณภาพพบว่าเมื่อข้อมูลมีค่า VIF เพิ่มขึ้นจะทำให้ค่าเฉลี่ยของ  $C$  เพิ่มขึ้น เมื่อจำนวนตัวแปรที่ใช้ทำนายเพิ่มขึ้นจะทำให้ค่าเฉลี่ยของ  $C$  เพิ่มขึ้น เมื่อพิจารณาที่แต่ละระดับของจำนวนตัวแปรที่ใช้ทำนายและค่า VIF พบว่าในกรณีที่ประชากรมีจำนวนตัวแปรเชิงคุณภาพเพิ่มขึ้นจะทำให้ค่าเฉลี่ยของ  $C$  ลดลง

#### 5.4 สรุปผลการเปรียบเทียบ

ในการศึกษาได้จำลองข้อมูลโดยกำหนดให้ตัวแปรตอบสนอง (ตัวแปรตาม) ซึ่งมี 2 ค่า มีความสัมพันธ์เชิงเส้นตรงกับตัวแปรที่ใช้ทำนาย โดยตัวแปรที่ใช้ทำนายมีทั้งตัวแปรเชิงคุณภาพและตัวแปรเชิงปริมาณ โดยตัวแปรเชิงปริมาณแต่ละตัวมีการแจกแจงปกติและความแปรปรวนไม่แตกต่างกัน แต่กำหนดให้มีตัวแปรที่ใช้ทำนายคู่ใดคู่หนึ่งมีความสัมพันธ์กันในตั้งแต่ระดับน้อย ( $VIF \leq 4$ ) ไปจนถึงระดับมาก ( $VIF > 10$ ) จากการเปรียบเทียบประสิทธิภาพการทำนายการเป็นสมาชิกกลุ่มที่ถูกต้องด้วยวิธีวิเคราะห์การถดถอยโลจิสติก วิธีวิเคราะห์จำแนกกลุ่มในเชิงเส้นตรง วิธีวิเคราะห์การถดถอยพหุคูณ และวิธีรีดจ์รีเกรสชันโดยใช้ค่าเฉลี่ยของความคลาดเคลื่อนในการจำแนกกลุ่ม ค่าเฉลี่ยของ B และค่าเฉลี่ยของ C พบว่าในทุกกรณีวิธีวิเคราะห์การถดถอยโลจิสติกและวิธีวิเคราะห์จำแนกกลุ่มในเชิงเส้นตรงมีประสิทธิภาพการทำนายการเป็นสมาชิกกลุ่มที่ถูกต้องไม่แตกต่างกัน และมีประสิทธิภาพในการจำแนกกลุ่มดีกว่าวิธีวิเคราะห์การถดถอยพหุคูณและวิธีรีดจ์รีเกรสชันถึงแม้จะตัวแปรที่ใช้ในการทำนายบางตัวจะมีความสัมพันธ์กันก็ตาม