

บทที่ 2

วิธีการดำเนินการศึกษา

2.1 วิธีการศึกษา

ในการศึกษาครั้งนี้ปริญญามีวิธีการดำเนินงานโดยสรุปดังนี้

1. ศึกษาตัวแปรทั้ง 45 ตัวที่ได้จากรายงานการศึกษาต้นแบบกลไกการบูรณาการงานพัฒนาสังคมของสำนักนโยบายและยุทธศาสตร์ สำนักงานปลัดกระทรวงการพัฒนาสังคมและความมั่นคงของมนุษย์
2. ทำการรวบรวมข้อมูลของตัวแปรต่างๆ ทั้ง 45 ตัวแปรจากข้อมูลปี 2548 จากแหล่งข้อมูลต่างๆที่หาได้ เช่น กรมการพัฒนาชุมชน กระทรวงมหาดไทย สำนักงานสถิติแห่งชาติ และ หน่วยงานอื่นๆ เป็นต้น
3. ทำการตรวจสอบความสมบูรณ์ของข้อมูลโดยเทคนิคการ cleaning data เพื่อให้ได้เพิ่มข้อมูลที่สมบูรณ์
4. ใช้เทคนิคการวิเคราะห์ข้อมูลทางสถิติของตัวแปรพหุที่เรียกว่า Cluster Analysis เพื่อทำการจำแนกข้อมูลออกเป็นกลุ่ม ๆ ตามวัตถุประสงค์ของโครงการ
5. นำข้อมูลของแต่ละกลุ่มมาพัฒนาตัวแบบเพื่อให้ได้สมการจำแนกกลุ่ม โดยใช้เทคนิคการวิเคราะห์ทางสถิติที่เรียกว่า Discriminant Analysis เพื่อใช้เป็นตัวแบบของข้อมูลแต่ละกลุ่ม
6. ใช้สมการจำแนกกลุ่มที่ได้จัดกลุ่มให้กับข้อมูลที่มาใหม่
7. นำตัวแบบที่พัฒนาได้ไปพัฒนาโปรแกรมคอมพิวเตอร์เพื่อประโยชน์ในการนำตัวแบบไปใช้ในการพยากรณ์สภาวะทางสังคมของประชาชน ซึ่งเป็นไปตามวัตถุประสงค์ของโครงการได้อย่างสะดวก

2.2. การดำเนินการศึกษา

ในการศึกษาครั้งนี้ปริญญได้ทำการศึกษาการสร้างตัวแบบจากกลุ่มตัวแปร 3 กลุ่มตามตัวแปรต้นแบบกลไกการบูรณาการงานพัฒนาสังคม ซึ่งประกอบด้วย 1) ตัวแปรความต้องการขั้นต่ำ 2) ตัวแปรความเสี่ยงของการพัฒนา และ 3) ตัวแปรที่แสดงศักยภาพสำหรับการพัฒนา ซึ่งมีรายละเอียดในการศึกษาการสร้างตัวแบบของตัวแปรแต่ละกลุ่มต่อไปนี้

ตัวแบบที่ 1 : ตัวแบบความต้องการขั้นต่ำ

จากรายงานการศึกษาต้นแบบกลไกการบูรณาการการพัฒนาสังคม ได้กำหนดตัวแปรที่ใช้พิจารณาความต้องการขั้นต่ำไว้ทั้งหมด 17 ตัวแปร ได้แก่

1. ร้อยละของหญิงตั้งครรภ์ที่ได้รับการดูแลก่อนคลอด (V1)
2. ร้อยละของหญิงคลอดลูกที่ได้รับการดูแลหลังคลอด (V2)
3. ร้อยละของเด็กอายุ 0 – 5 ปี ได้กินอาหารเพียงพอ (V3)
4. ร้อยละของเด็กอายุ 6 – 15 ปี ได้กินอาหารอย่างถูกต้อง (V4)
5. ร้อยละของจำนวนคนในครัวเรือน ได้กินอาหารที่มีคุณภาพและได้มาตรฐาน (V5)
6. ร้อยละของเด็กอายุ 0 – 1 ปี ได้วัคซีนครบ (V6)
7. ร้อยละของเด็กอายุ 6 – 12 ปี ได้รับการฉีดวัคซีนครบ (V7)
8. จำนวนครัวเรือนที่มีความรู้ในการใช้ยาที่ถูกต้อง (V8)
9. ร้อยละของคนอายุ 35 ปีขึ้นไปที่ได้รับการตรวจสุขภาพประจำปี (V9)
10. ร้อยละของผู้มีงานทำ (V10)
11. จำนวนคนผู้สูงอายุที่มีงานทำ (V11)
12. จำนวนครัวเรือนที่ไม่ถูกรบกวนจากมลพิษ (V12)
13. จำนวนพื้นที่ปลูกป่าและไม้ยืนต้น (ไร่ / คน) (V13)*
14. ร้อยละของครัวเรือนที่มีน้ำใช้เพียงพอตลอดปี (V14)
15. จำนวนประชากรที่ได้เรียนหนังสือ (คน) (V15)*
16. มูลค่าของเบี้ยประกันสุขภาพส่วนบุคคล (บาท) (V16)
17. จำนวนคนพิการที่กู้เงินกองทุนฟื้นฟู (คน) (V17)

(* ตัวแปรที่เปลี่ยนไปจากตัวแปรเดิม)

จากข้อมูลทั้งหมด 69,192 หมู่บ้าน **พบ**การตรวจสอบความสมบูรณ์ของข้อมูลเบื้องต้นพบว่ามี 23,165 หมู่บ้าน มีข้อมูลไม่สมบูรณ์ หรือข้อมูลของบางตัวแปรสูญหาย และมี 532 หมู่บ้าน ที่ข้อมูลของบางตัวแปรไม่สมเหตุสมผล เช่น ข้อมูลเกี่ยวกับร้อยละมีรายงานข้อมูลมากกว่า 100 ซึ่งเป็นค่าที่เป็นไปไม่ได้ เป็นต้น เนื่องจากข้อมูลมีเป็นจำนวนมาก การคัดข้อมูลที่ไม่สมบูรณ์ทั้งจึงไม่ส่งผล

Deleted: จำนวนเป็นข้อมูลจากภาคเหนือ 15,175 หมู่บ้าน ภาคตะวันออกเฉียงเหนือ 30,873 หมู่บ้าน ภาคกลาง 14,911 หมู่บ้าน และภาคใต้ 8,233 หมู่บ้าน จาก

กระทบต่อการประมวลผลมากนัก ดังนั้นเพื่อให้การประมวลผลมีความถูกต้องและไม่เป็นปัญหาต่อการประมวลผล จึงคัดข้อมูลที่ไม่สมบูรณ์เหล่านั้นทิ้ง ทำให้เหลือข้อมูลเพื่อใช้ประมวลผลทั้งสิ้น 45,495 หมู่บ้าน

ในการสร้างและพัฒนาตัวแบบการจำแนกหรือตัวแบบการพยากรณ์ใดๆ โดยใช้ข้อมูลจากเพิ่มข้อมูลที่มีอยู่ การพิจารณาความเป็นไปได้และลักษณะการกระจายของข้อมูลรายตัวแปรถือเป็นสิ่งสำคัญ อีกทั้งเทคนิคทางสถิติโดยทั่วไปจะมีข้อกำหนดเกี่ยวกับลักษณะการกระจายของตัวแปรไว้ ดังนั้นเพื่อให้ได้ตัวแบบที่สมเหตุผลและมีประโยชน์ในทางปฏิบัติ จึงต้องพิจารณาลักษณะการกระจายของตัวแปรทั้ง 17 ตัว ก่อนที่จะนำมาใช้สร้างตัวแบบสำหรับข้อมูลความต้องการขั้นต่ำ ซึ่งจากการศึกษาพบว่า

1) ตัวแปร V11 และ ตัวแปร V16 มีค่าเท่ากันทั้งหมด เนื่องจากว่าข้อมูลที่ได้เป็นข้อมูลในระดับประเทศ เมื่อนำมาคิดเป็นรายหมู่บ้าน จึงต้องมีการนำมาเฉลี่ย จึงทำให้มีผลเป็นค่าที่เท่ากันทั้งหมด เมื่อค่าของตัวแปรมีค่าเท่ากันทั้งหมด ตัวแปรทั้งสองจึงไม่อาจสามารถใช้เป็นตัวแปรที่จำแนกความแตกต่างของความต้องการขั้นต่ำได้ จึงทำให้ต้องตัดตัวแปรเหล่านี้ทิ้งไป

2) ข้อมูลตัวแปรที่มีค่าบางค่ามีจำนวนมากผิดปกติ ได้แก่

ตัวแปร V1 มีค่าเท่ากับ 100 จำนวน 45,025 หมู่บ้าน

ตัวแปร V2 มีค่าเท่ากับ 100 จำนวน 44,960 หมู่บ้าน

ตัวแปร V3 มีค่าเท่ากับ 100 จำนวน 43,710 หมู่บ้าน

ตัวแปร V4 มีค่าเท่ากับ 100 จำนวน 42,983 หมู่บ้าน

ตัวแปร V6 มีค่าเท่ากับ 100 จำนวน 44,926 หมู่บ้าน

ตัวแปร V7 มีค่าเท่ากับ 100 จำนวน 43,083 หมู่บ้าน

ตัวแปร V13 มีค่าเท่ากับ 0 จำนวน 36,455 หมู่บ้าน

เนื่องจากค่าส่วนใหญ่ของตัวแปรเหล่านี้เป็นไปดั่งข้างต้น จึงทำให้ค่าอื่นๆของตัวแปรดังกล่าวถือเป็นค่าผิดปกติ เช่น ตัวแปร V1 ร้อยละของหญิงตั้งครรภ์ได้รับการดูแลก่อนคลอด มีค่าเท่ากับ 100 จำนวน 45,025 หมู่บ้าน ดังนั้น V1 จะมีค่าอื่นๆที่ต่ำกว่า 100 อยู่จำนวน 470 หมู่บ้าน ซึ่งค่าอื่นๆของ V1 ที่มีจำนวน 470 หมู่บ้านนี้ ถือเป็นค่าผิดปกติของตัวแปร V1 เป็นต้น และค่าผิดปกติของตัวแปรใดๆจะมีอิทธิพลทำให้ค่าสถิติของตัวแปรเหล่านั้นผิดจากความเป็นจริง ในทางปฏิบัติจะต้องมีการตรวจสอบว่าสาเหตุของค่าผิดปกติเหล่านี้มาจากอะไร แล้วจึงทำการแก้ไข หรือถ้ามีข้อมูลจำนวนมากก็สามารถตัดข้อมูลรายหมู่บ้านที่เป็นข้อมูลผิดปกติของตัวแปรเหล่านั้นทิ้งได้ เช่น ตัดข้อมูล 470 หมู่บ้านของตัวแปร V1 ทิ้ง เป็นต้น สำหรับกรณีของการตัดข้อมูลผิดปกติทิ้ง จะทำให้ข้อมูลที่เหลือมีค่าเท่ากันทั้งหมด เช่น ตัวแปร V1 ถ้าตัดข้อมูลผิดปกติ 470 หมู่บ้านทิ้ง จะทำให้เหลือข้อมูลของตัวแปร V1 มีค่าเท่ากับ 100 เท่ากันหมดจำนวน 45,025 หมู่บ้าน ตัวแปร V1 จึงเป็นตัวแปรที่สอดคล้องกับข้อ 1.) ดังนั้นจึงทำให้ไม่สามารถนำตัวแปร V1 มาพิจารณาได้ ในทำนองเดียวกันกับตัวแปร V2,V3,V4,V6,V7,V13 ที่ต้องตัดออกการพิจารณาใช้สร้างตัวแบบ

3) ข้อมูลตัวแปร V17 มีค่า 2 ค่า คือมีค่าเป็น 1 จำนวน 44,934 ค่า และมีค่าเป็น 2 จำนวน 556 ค่า ซึ่งมีลักษณะข้อมูลเช่นเดียวกับข้อ 2.) จึงทำให้ต้องตัดตัวแปรนี้ออกจากการใช้พิจารณาเช่นกัน

4) สำหรับข้อมูลตัวแปร V5,V8,V9,V10,V12,V14 และ V15 ในแต่ละตัวแปรมีรายการข้อมูลที่ผิดปกติอยู่ มีจำนวนแตกต่างกันไป ไม่มีจำนวนมากเหมือนตัวแปรในข้อ 2.) จึงทำให้ต้องพิจารณาการตัดข้อมูลที่ผิดปกติบาง

หมู่บ้านออก เพื่อให้ได้ข้อมูลของทุกตัวแปรที่เหลือ มีค่าผิดปกติเล็กน้อยที่สุดหรือไม่มีเลย จากการพิจารณารายตัวแปรไปพร้อมๆกันทุกตัวแปร พบว่าสามารถกำจัดค่าผิดปกติของตัวแปร V5,V8,V9,V10,V12,V14 และ V15 ได้เกือบทั้งหมด

5) สรุปจำนวนตัวแปรที่นำมาจัดกลุ่มมี 7 ตัว ได้แก่ V5,V8,V9,V10,V12,V14 และ V15 จำนวน 1,363 หมู่บ้าน

การพัฒนาตัวแบบเพื่อการจำแนกกลุ่ม

ในการวิเคราะห์ข้อมูลเพื่อพัฒนาตัวแบบการจำแนกกลุ่ม โดยใช้ข้อมูล 1,363 หมู่บ้าน 7 ตัวแปร โดยใช้เทคนิคการวิเคราะห์แบบ Discriminant Analysis นั้น มีความจำเป็นที่จะต้องทราบกลุ่มของแต่ละหมู่บ้านเสียก่อน แต่เนื่องจากข้อมูลที่มีอยู่ยังไม่ทราบกลุ่มของแต่ละหมู่บ้าน ซึ่งในความหมายของทางกระทรวงการพัฒนาสังคมและความมั่นคงของมนุษย์ได้แบ่งหมู่บ้านตามความต้องการขั้นต่ำออกเป็น 3 กลุ่ม ได้แก่

กลุ่มหมู่บ้านที่**ต้องการ**ความต้องการขั้นต่ำ**น้อย** กำหนดให้เป็นกลุ่มไฟเขียว

กลุ่มหมู่บ้านที่**ต้องการ**ความต้องการขั้นต่ำปานกลาง กำหนดให้เป็นกลุ่มไฟเหลือง

กลุ่มหมู่บ้านที่**ต้องการ**ความต้องการขั้นต่ำมาก กำหนดให้เป็นกลุ่มไฟแดง

Deleted: ได้รับ

Deleted: สูง

Deleted: ได้รับ

Deleted: ได้รับ

Deleted: น้อย

ดังนั้นจึงต้องมีการวิเคราะห์เพื่อกำหนดกลุ่มให้กับแต่ละหมู่บ้านเสียก่อน ซึ่งในการนี้คณะที่ปรึกษาได้ใช้เทคนิคการวิเคราะห์การจัดกลุ่มด้วยเทคนิคที่เรียกว่า Cluster Analysis โดยในขั้นแรกจะทำการหาจุดศูนย์กลางรวมของแต่ละกลุ่มสมมติ 3 กลุ่ม จากนั้นจะคำนวณหาระยะห่างของข้อมูลแต่ละหมู่บ้านกับจุดศูนย์กลางรวมของแต่ละกลุ่ม ถ้าข้อมูลของหมู่บ้านใดใกล้จุดศูนย์กลางของกลุ่มใด ก็จะจัดหมู่บ้านนั้นเข้ากับกลุ่มนั้น แล้วปรับจุดศูนย์กลางของทั้งกลุ่มที่หมู่บ้านใหม่เข้าไปและจุดศูนย์กลางของกลุ่มที่หมู่บ้านใหม่ออกมา ทำเช่นนี้ไปเรื่อยๆ กับทุกๆ หมู่บ้าน และหลายๆรอบจนกระทั่งไม่มีระยะห่างเปลี่ยนแปลง ระยะห่างคำนวณโดยวิธีที่เรียกว่า Euclidean distance ซึ่งเป็นเทคนิคที่จะขจัดปัญหาเกี่ยวกับความสัมพันธ์ในระหว่างตัวแปร(Multicollinearity) ของตัวแปรที่ใช้ในการจัดกลุ่ม มีสูตรในการคำนวณ ดังนี้

$$d = \sqrt{\sum_j (X_{ij} - X_{1j})^2}$$

เมื่อ x_{ij} เป็นค่าข้อมูลของรายการที่ i ; $i = 1, 2, \dots, 1363$ ของตัวแปรที่ j ; $j = 5, 8, 9, 10, 12, 14, 15$

และจากข้อมูลชุดนี้ทั้งหมด คณะที่ปรึกษาได้ทำการประมวลผลการวิเคราะห์จัดกลุ่ม โดยโปรแกรมสำเร็จรูปทางสถิติได้ผลลัพธ์การจัดกลุ่มในรอบสุดท้ายของการจัดกลุ่มดังตารางที่ 1.1 และ 1.2 ในภาคผนวก

จากตารางที่ 1.1 และ 1.2 ในภาคผนวกและคำแนะนำของผู้เชี่ยวชาญทางด้านสังคมศาสตร์และความมั่นคงของมนุษย์ จัดกลุ่มได้เป็น

กลุ่มที่ 1 เป็นกลุ่มไฟเขียว มีหมู่บ้านที่จัดอยู่ในกลุ่มอยู่จำนวนทั้งหมด

572 หมู่บ้าน

Deleted: 464

กลุ่มที่ 2 เป็นกลุ่มไฟเหลือง มีหมู่บ้านที่จัดอยู่ในกลุ่มอยู่จำนวนทั้งหมด

327 หมู่บ้าน

กลุ่มที่ 3 เป็นกลุ่มไฟแดง มีหมู่บ้านที่จัดอยู่ในกลุ่มอยู่จำนวนทั้งหมด

464 หมู่บ้าน

Deleted: 572

ตารางที่ 1 จำนวนหมู่บ้านจำแนกตามประเภทของกลุ่มและภาคต่างๆ ของตัวแบบที่ 1

กลุ่ม	ภาค				รวม (%)
	เหนือ (%)	กลาง (%)	ตะวันออกเฉียงเหนือ (%)	ใต้ (%)	
ไฟเขียว	157 (11.5)	107 (7.9)	226 (16.6)	82 (6.0)	572 (42.0)
ไฟเหลือง	109 (8.0)	53 (3.9)	101 (7.4)	64 (4.7)	327 (24.0)
ไฟแดง	124 (9.1)	78 (5.7)	159 (11.7)	103 (7.6)	464 (34.0)
รวม	390 (28.6)	238 (17.5)	486 (35.7)	249 (18.3)	1363 (100.0)

จากตารางที่ 1 ข้างต้น พบว่าหมู่บ้านที่จัดอยู่ในกลุ่มไฟแดง อยู่ในภาคตะวันออกเฉียงเหนือมีจำนวนสูงสุด (16.6%) รองลงมาได้แก่เป็นหมู่บ้านในภาคเหนือ ภาคกลางและภาคใต้ ตามลำดับ สำหรับหมู่บ้านที่จัดเป็นไฟเขียว และไฟเหลืองมีลักษณะการกระจายของหมู่บ้านรูปแบบเดียวกันคือ ภาคตะวันออกเฉียงเหนือ ภาคเหนือ ภาคใต้ และภาคกลาง ตามลำดับ

ในการสร้างตัวแบบเพื่อการจำแนกกลุ่ม โดยใช้เทคนิคการวิเคราะห์การจำแนกกลุ่ม (Discriminant Analysis) คณะที่ปรึกษาใช้หลักการของ Fisher's method เพื่อหาสมการจำแนกกลุ่มหรือเรียกว่า Fisher Discriminant Function ซึ่งคิดค้นโดย R.A.Fisher ในปี ค.ศ. 1936 โดยมีตัวแบบสมการเชิงเส้นเป็น

$$D = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon \dots \dots \dots (1)$$

Deleted: 24

Deleted: 78

Deleted: 159

Deleted: 103

Deleted: 464

Deleted: 9.1

Deleted: 5.7

Deleted: 11.7

Deleted: 7.6

Deleted: 3

Deleted: 57

Deleted: 10

Deleted: 226

Deleted: 82

Deleted: 572

Deleted: 11.5

Deleted: 7.9

Deleted: 6.6

Deleted: 6.0

Deleted: 2

เมื่อ D เป็นตัวแปรตามซึ่งเป็นตัวแปรที่มีค่าข้อมูลเป็นข้อมูลเชิงกลุ่ม ซึ่งในที่นี้ D จะมีค่า 3 ค่า คือ

$$D = \begin{cases} 1 & \text{ถ้าหมู่บ้านจัดอยู่ในกลุ่มไฟเขียว} \\ 2 & \text{ถ้าหมู่บ้านจัดอยู่ในกลุ่มไฟเหลือง} \\ 3 & \text{ถ้าหมู่บ้านจัดอยู่ในกลุ่มไฟแดง} \end{cases}$$

และ $X_1, X_2, \dots, X_p$ เป็นตัวแปรอิสระหรือตัวแปรจำแนกกลุ่มซึ่งเป็นตัวแปรเชิงปริมาณ

$\beta_0, \beta_1, \dots, \beta_p$ แทนสัมประสิทธิ์ของสมการจำแนกกลุ่ม

p แทนจำนวนตัวแปรจำแนกกลุ่ม

โดยมีเงื่อนไขของการสร้างตัวแบบคือ

1. ตัวแปรอิสระ p ตัว ($X_1, X_2, \dots, X_p$) ของข้อมูลทั้ง 3 กลุ่มไม่จำเป็นต้องมีการแจกแจงปกติแบบพหุ (multivariate normal distribution)
2. เมทริกซ์ของความแปรปรวนขนาด $p \times p$ ของแต่ละกลุ่มต้องเท่ากัน

สำหรับการตรวจสอบเงื่อนไขการสร้างตัวแบบเพื่อการจำแนกกลุ่ม ขณะที่ปริกษาได้ทำการศึกษาความผิดปกติและการแจกแจงปกติของตัวแปรแต่ละตัวของข้อมูลในแต่ละกลุ่มไฟสี พร้อมกันนั้นได้ตรวจสอบความผิดปกติร่วมของตัวแปรของข้อมูลในแต่ละกลุ่มไฟสี โดยใช้สถิติ Mahalanobis distance ที่ค่านัยสำคัญ .001 ของการแจกแจงแบบโคกำลัง ที่ระดับองศาของความอิสระเท่ากับจำนวนตัวแปรที่ศึกษา จากการศึกษาตัวแปรอิสระดังกล่าวในแต่ละกลุ่มไฟสี สรุปได้ว่าตัวแปรอิสระส่วนใหญ่มีการแจกแจงแบบปกติและตัวแปรอิสระบางตัวพออนุโลมให้มีการแจกแจงแบบปกติได้ทั้ง 3 กลุ่มข้อมูล ความผิดปกติอยู่ในภาวะที่สามารถรับได้ ผลการตรวจสอบการแจกแจงแบบปกติและความผิดปกติของตัวแปรแต่ละตัวในแต่ละกลุ่มปรากฏในรูปที่ 1.1-1.42 ในภาคผนวก สำหรับกรณีของความผิดปกติร่วม ปรากฏว่ามีข้อมูลผิดปกติร่วมในกลุ่มไฟเขียว 5 หมู่บ้าน กลุ่มไฟเหลือง 4 หมู่บ้าน และกลุ่มไฟแดงไป 11 หมู่บ้าน จึงต้องตัดข้อมูลหมู่บ้านดังกล่าวทิ้ง ดังนั้นจึงเหลือข้อมูลที่จะใช้สำหรับการสร้างตัวแบบ Fisher Discriminant Function ในกลุ่มไฟเขียว 459 หมู่บ้าน ในกลุ่มไฟเหลือง 323 หมู่บ้าน และในกลุ่มไฟแดง 561 หมู่บ้าน รวมทั้งสิ้น 1,343 หมู่บ้าน

สำหรับการตรวจสอบเงื่อนไขการเท่ากันของความผันแปรของตัวแปรในแต่ละกลุ่มพิจารณาได้จากรูปภาพที่ 1.43-1.45 ในภาคผนวก ซึ่งพบว่าการกระจายของคะแนนที่ได้จากตัวแบบของฟังก์ชันการจำแนกกลุ่มที่ได้จากแต่ละกลุ่มไฟเขียว ไฟเหลือง และไฟแดง มีการกระจายที่พออนุโลมว่าไม่มีความแตกต่างกันมากนัก และนำไปสู่การสร้างเงื่อนไขของการเท่ากันของความผันแปรของตัวแปรในแต่ละกลุ่ม

นอกจากนั้นยังต้องมีการตรวจสอบความสัมพันธ์ในระหว่างตัวแปร V5,V8,V9,V10,V12,V14 และ V15 เพื่อตรวจสอบการเกิด Multicollinearity ซึ่งถ้าเกิด Multicollinearity ระหว่างตัวแปรอิสระขึ้นจะมีผลทำให้ค่าประมาณของสัมประสิทธิ์ของตัวแปรเหล่านี้ในตัวแบบมีค่าคลาดเคลื่อนได้ ซึ่งจะส่งผลต่อการพยากรณ์ ขณะที่ปรึกษาได้พิจารณาความสัมพันธ์ระหว่างตัวแปรอิสระเหล่านี้จากสถิติสัมประสิทธิ์สหสัมพันธ์ ปรากฏในตารางที่ 1.3 ในภาคผนวก พบว่า V8 มีความสัมพันธ์กับ V12 ในระดับที่สูง คือ .948 และมีความสัมพันธ์กับ V15 ในระดับ .697 ซึ่งไม่สูงมากนัก ในขณะที่ V12 มีความสัมพันธ์กับ V15 ในระดับ .686 ซึ่งไม่สูงมากเช่นกัน ดังนั้นเพื่อไม่ให้เกิดในกรณีเรียกว่า Multicollinearity ระหว่างตัวแปร จึงตัดแปร V8 ออกจากการพิจารณา เพราะถือว่าตัวแปร V12 ทำหน้าที่แทนแล้ว

โดยสรุปในการสร้างตัวแบบจะใช้ข้อมูลจากตัวแปรทั้งสิ้น 6 ตัวแปร ได้แก่ V5,V9,V10,V12, V14 และ V15 และมีรายการข้อมูล 1,343 หมู่บ้าน

จากการประมวลผลด้วยโปรแกรมสำเร็จรูปทางสถิติโดยใช้เทคนิคการจำแนกกลุ่ม Discriminant Analysis พัฒนาและเลือกตัวแปรเข้าไปในตัวแบบ ด้วยเทคนิคทางสถิติที่เรียกว่า Stepwise method ได้ผลลัพธ์สุดท้าย ปรากฏในตารางที่ 1.4-1.7 ในภาคผนวก จากตารางที่ 1.4 พบว่าค่าเฉลี่ยของตัวแปรอิสระแต่ละตัวมีความแตกต่างกันเมื่ออยู่ต่างกลุ่มกัน ตารางที่ 1.5 แสดงให้เห็นว่าค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรอิสระคู่ใดมีค่าค่อนข้างต่ำ จึงทำให้เชื่อได้ว่าจะไม่เกิดปัญหาที่เรียกว่า Multicollinearity ในตัวแบบ ตารางที่ 1.6 เป็นค่าสัมประสิทธิ์ของตัวแปรในสมการจำแนกกลุ่ม ซึ่งจากตารางนี้ทำให้ได้ตัวแบบของการจำแนกกลุ่มเป็นดังนี้

1. ตัวแบบการแบ่งกลุ่มของกลุ่มไฟเขียว คือ

$$D = -4638.603 + 43.73 (V5) + 2.966 (V9) + 17.546 (V10) - 0.076 (V12) + 30.474 (V14) + 0.099(V15)$$

2. ตัวแบบสำหรับการแบ่งกลุ่มของกลุ่มไฟเหลือง คือ

$$D = -4484.954 + 43.172(V5) + 2.718(V9) + 16.822(V10) - 0.12(V12) + 30.485(V14) + 0.084(V15)$$

3. ตัวแบบสำหรับการแบ่งกลุ่มของกลุ่มไฟแดง คือ

$$D = -4592.584 + 43.529(V5) + 2.949(V9) + 17.691(V10) - 0.143(V12) + 30.21(V14) + 0.073(V15)$$

Deleted: $D = -4638.603 + 43.73 (V5) + 2.966 (V9) + 17.546 (V10) - 0.076 (V12) + 30.474 (V14) + 0.099(V15)$

Deleted: $D = -4592.584 + 43.529(V5) + 2.949(V9) + 17.691(V10) - 0.143(V12) + 30.21(V14) + 0.073(V15)$

แนวคิดการใช้ตัวแบบในการพยากรณ์กลุ่ม

สำหรับการใช้ตัวแบบที่สร้างขึ้นเพื่อจัดกลุ่ม(พยากรณ์กลุ่ม)ให้แก่หมู่บ้านใหม่ เช่น จะใช้ตัวแบบเพื่อพยากรณ์ว่าหมู่บ้านที่ได้ข้อมูลมาใหม่ควรจะอยู่ในกลุ่มหมู่บ้านที่มีระดับความต้องการขั้นต่ำอยู่ในระดับใด เช่น ไฟเขียว ไฟเหลือง หรือไฟแดง นั้น จะใช้หลักเกณฑ์ของ Fisher โดยใช้สมการ Linear classification function ที่มีแนวความคิดว่าระยะห่างระหว่างข้อมูลของตัวแปรของหมู่บ้านใหม่ห่างจุดศูนย์กลางของกลุ่มข้อมูลกลุ่มใดน้อยที่สุด ก็จะจัดหมู่บ้านใหม่ให้กับกลุ่มนั้น ซึ่งระยะห่างกำลังสองคำนวณได้จาก

$$D_i^2(X) = (X - \bar{X}_i)'S^{-1}(X - \bar{X}_i) \dots \dots \dots (2)$$

เมื่อ S เป็นเมทริกซ์ความแปรปรวนร่วมของข้อมูลทุกกลุ่ม และ $\bar{X}_i$ เป็นค่าเฉลี่ยของข้อมูลกลุ่มที่ i และจะจัดข้อมูลหมู่บ้านใหม่ให้กับกลุ่มที่ i ถ้า $D_i^2(X)$ มีค่าน้อยที่สุด จากสมการที่ (2) สามารถพัฒนาให้เป็นสมการที่ (1) ได้โดยวิธีทางคณิตศาสตร์ ซึ่งจะช่วยให้เงื่อนไขการจัดหมู่บ้านใหม่เปลี่ยนไป โดยจะจัดหมู่บ้านใหม่ที่มีข้อมูลมาใหม่ให้กับกลุ่มที่ i ถ้า D (Discriminant scores) ของกลุ่มใดมีค่ามากที่สุด โดยการแทนค่าของตัวแปรอิสระหรือตัวแปรจำแนกกลุ่มของหมู่บ้านใหม่ลงในสมการจำแนกกลุ่มแล้วคำนวณหาค่าตัวแปรตาม D ถ้าค่า D ของกลุ่มใดมีค่ามากที่สุดก็จะจัดหมู่บ้านใหม่นั้นให้อยู่กับกลุ่มนั้น

ตารางที่ 1.7 ในภาคผนวก เป็นผลการพยากรณ์ข้อมูลรายหมู่บ้านตามความต้องการขั้นต่ำ แล้วนำไปจัดเป็นหมู่บ้านไฟเขียว ไฟเหลือง หรือ ไฟแดง โดยวิธีพยากรณ์ข้อมูลทั้งหมด(Original) และวิธีพยากรณ์ข้อมูลบางส่วน(Cross-Validated) ผลปรากฏว่าทั้งสองวิธีให้ผลการพยากรณ์หมู่บ้านได้ถูกต้องใกล้เคียงกัน คือประมาณ 94% เท่ากัน จึงสามารถกล่าวได้ว่าผลการพยากรณ์ของตัวแบบที่สร้างขึ้นอยู่ในระดับที่น่าพอใจ

ตัวแบบที่ 2 : ตัวแบบความเสี่ยงของการพัฒนา

ในทำนองเดียวกันกับการศึกษาตัวแบบที่1 จากรายงานการศึกษาต้นแบบกลไกการบูรณาการพัฒนาสังคม ได้กำหนดตัวแปรที่ใช้พิจารณาความเสี่ยงของการพัฒนาไว้ทั้งหมด 13 ตัวแปร ได้แก่

1. จำนวนผู้ป่วยด้วยโรคพื้นฐานจากการกำหนดในฐานข้อมูลของ กชช. 2ค เช่น โรคคอติด ไทรอยด์ ภาวะตะกั่ว โลหิต และอื่นๆ เป็นต้น (V18)
2. จำนวนผู้ได้รับการบำบัดยาเสพติด (V19)

3. จำนวนผู้สูงอายุที่ถูกทอดทิ้ง (V20)
 4. อัตราการหย่าร้างต่อการสมรส (V21)
 5. จำนวนเด็กที่ถูกทอดทิ้ง (V22)
 6. จำนวนคดีเด็กและเยาวชนที่รับใหม่ในศาล (V23)*
 - 7. ปริมาณขยะที่เกิดขึ้นต่อวัน(หน่วย:ตันต่อวัน) (V24)
 8. ค่าใช้จ่ายในการซื้อปุ๋ยต่อปี (V25)*
 9. ค่าใช้จ่ายในการซื้อสารเคมีป้องกันและกำจัดแมลง วัชพืชต่อปี (V26)*
 10. มูลค่าการนำเข้าสินค้าอุปโภคและบริโภค (V27)*
 11. ปริมาณการใช้พลังงานไฟฟ้าเชิงพาณิชย์ต่อปี(หน่วยเป็น กิโลวัตต์) (V28)
 12. จำนวนคนที่ถูกจับกุมกรณีค้าประเวณี หรือยาเสพติด (V29)
 13. จำนวนคดีที่เกี่ยวข้องกับชีวิต ร่างกาย และเพศ ที่จับกุมได้ (V30)*
- (* ตัวแปรที่เปลี่ยนไปจากตัวแปรเดิม)

ในการตรวจสอบความสมบูรณ์ของข้อมูลเบื้องต้นสำหรับตัวแบบที่ 2 พบว่ามีจำนวนข้อมูล 130 หมู่บ้าน จาก 69,192 หมู่บ้าน มีข้อมูลไม่สมบูรณ์ หรือข้อมูลของบางตัวแปรสูญหาย จึงตัดข้อมูลที่ไม่สมบูรณ์เหล่านั้นทิ้ง ทำให้เหลือข้อมูลเพื่อใช้พิจารณาทั้งสิ้น 69,062 หมู่บ้าน

ในการศึกษาการกระจายของตัวแปรทั้ง 13 ตัวที่นำมาใช้สร้างตัวแบบสำหรับข้อมูลความเสี่ยงของการพัฒนาพบว่า

1) ตัวแปร V21 และ ตัวแปร V27 มีค่าเท่ากันทั้งหมด เนื่องจากข้อมูลที่ได้เป็นข้อมูลระดับประเทศ ดังนั้นตัวแปรดังกล่าวจึงไม่สามารถใช้เป็นตัวแปรจำแนกความเสี่ยงของการพัฒนาได้ จึงทำให้ต้องตัดตัวแปรเหล่านี้ทิ้ง

- 2) ข้อมูลตัวแปรที่มีค่าบางค่า มีจำนวนมากผิดปกติ ได้แก่
- ตัวแปร V18 มีค่าเท่ากับ 0 จำนวน 51,565 หมู่บ้าน
 - ตัวแปร V19 มีค่าเท่ากับ 0 จำนวน 55,063 หมู่บ้าน
 - ตัวแปร V20 มีค่าเท่ากับ 0 จำนวน 65,156 หมู่บ้าน
 - ตัวแปร V22 มีค่าเท่ากับ 0 จำนวน 68,268 หมู่บ้าน
 - ตัวแปร V23 มีค่าเท่ากับ 0 จำนวน 54,592 หมู่บ้าน
 - ตัวแปร V30 มีค่าเท่ากับ 0 จำนวน 66,551 หมู่บ้าน

ในทำนองเดียวกันกับการพิจารณาตัวแปรในตัวแบบที่ 1 เนื่องจากตัวแปร V19,V20,V22,V23 และ V30 มีค่าเป็น 0 ซ้ำกันเป็นจำนวนมาก จึงต้องตัดตัวแปรเหล่านี้ออกจากการพิจารณา สำหรับตัวแปร V18 มีจำนวนค่าข้อมูลที่ซ้ำกันน้อยกว่า 80% ของข้อมูลทั้งหมด จึงพยายามที่จะใช้ข้อมูลนี้ต่อไป

3) สำหรับข้อมูลตัวแปร V24,V25,V26,V28, และ V29 ในแต่ละตัวแปรมีค่าข้อมูลที่ผิดปกติอยู่จำนวนแตกต่างกันไป มีจำนวนไม่มากเหมือนตัวแปรในข้อ 2.) จึงทำให้ต้องพิจารณาตัดข้อมูลที่ผิดปกติบางหมู่บ้านของแต่ละตัวแปรที่ละตัวออก เพื่อให้ได้ข้อมูลของทุกตัวแปรที่เหลือ มีค่าผิดปกติเหลือน้อยที่สุดหรือไม่มีเลย จากการพิจารณารายตัวแปรไปพร้อมๆกันทุกตัวแปร พบว่าสามารถกำจัดค่าผิดปกติของตัวแปร V24,V25,V26,V28, และ V29 ได้เกือบทั้งหมด เหลือข้อมูลที่ใช้พิจารณาสร้างตัวแบบทั้งสิ้น 36,089 หมู่บ้าน ซึ่งในจำนวนนี้ไม่สามารถกำจัดค่าผิดปกติของตัวแปร V18 ได้ จึงมีความจำเป็นต้องตัดตัวแปร V18 ทิ้ง

4) สรุปมีตัวแปรที่ใช้พิจารณาจัดกลุ่มเพียง 5 ตัวแปรคือ V24,V25,V26,V28, และ V29 และมีข้อมูลจำนวนทั้งสิ้น 36,089 หมู่บ้าน

การพัฒนาตัวแบบเพื่อการจำแนกกลุ่ม

ในทำนองเดียวกันกับการพัฒนาตัวแบบที่ 1 การพัฒนาตัวแบบที่ 2 จึงมีความจำเป็นต้องมีการวิเคราะห์เพื่อการจัดกลุ่มข้อมูลเสียก่อน ในกรณีนี้ขณะที่ปรึกษาโดยกระทรวงการพัฒนาสังคมและความมั่นคงของมนุษย์ได้แบ่งหมู่บ้านตามความเสี่ยงของการพัฒนาออกเป็น 3 กลุ่ม ได้แก่

กลุ่มหมู่บ้านที่มีความเสี่ยงต่อการพัฒนาต่ำ กำหนดให้เป็นกลุ่มไฟเขียว

กลุ่มหมู่บ้านที่มีความเสี่ยงต่อการพัฒนาปานกลาง กำหนดให้เป็นกลุ่มไฟเหลือง

กลุ่มหมู่บ้านที่มีความเสี่ยงต่อการพัฒนามาก กำหนดให้เป็นกลุ่มไฟแดง

Deleted: ๑

๑

๑

-

จากข้อมูล 5 ตัวแปร จำนวน 36,089 หมู่บ้าน ขณะที่ปรึกษาทำการวิเคราะห์เพื่อการจัดกลุ่มด้วยเทคนิคทางสถิติ Cluster Analysis ได้ผลลัพธ์ปรากฏในตารางที่ 2.1 และ 2.2 ในภาคผนวก

จากตารางที่ 2.1 และ 2.2 และคำแนะนำของผู้เชี่ยวชาญทางด้านสังคมศาสตร์และความมั่นคงของมนุษย์ จัดกลุ่มหมู่บ้านได้เป็น

กลุ่มที่ 1	กลุ่มไฟเขียว	มีหมู่บ้านที่จัดอยู่ในกลุ่มนี้จำนวนทั้งหมด	16,424	หมู่บ้าน
กลุ่มที่ 2	กลุ่มไฟเหลือง	มีหมู่บ้านที่จัดอยู่ในกลุ่มนี้จำนวนทั้งหมด	11,323	หมู่บ้าน
กลุ่มที่ 3	กลุ่มไฟแดง	มีหมู่บ้านที่จัดอยู่ในกลุ่มนี้จำนวนทั้งหมด	8,342	หมู่บ้าน

ตารางที่ 2 จำนวนหมู่บ้านจำแนกตามประเภทของกลุ่มและภาคต่างๆของตัวแบบที่ 2

กลุ่ม	ภาค				รวม (%)
	เหนือ (%)	กลาง (%)	ตะวันออกเฉียงเหนือ (%)	ใต้ (%)	
ไฟเขียว	2598 (7.2)	105 (0.3)	13473 (37.3)	248 (0.7)	16424 (45.5)
ไฟเหลือง	3207 (8.9)	3552 (9.8)	2497 (6.9)	2067 (5.7)	11323 (31.4)
ไฟแดง	1859 (5.2)	452 (1.3)	5572 (15.4)	459 (1.3)	8342 (23.1)
รวม	7664 (21.2)	4109 (11.4)	21542 (59.7)	2774 (7.7)	36089 (100.0)

จากตารางที่ 2 ชำ้ค้นพบว่าหมู่บ้านที่จัดอยู่ในกลุ่มไฟแดง อยู่ในภาคตะวันออกเฉียงเหนือสูงสุด (15.4%) รองลงมาได้แก่เป็นหมู่บ้านในภาคเหนือ ภาคกลางและภาคใต้ ใกล้เคียงกัน สำหรับหมู่บ้านที่จัดเป็นไฟเขียวอยู่ในภาคตะวันออกเฉียงเหนือสูงสุด (37.3%) รองลงมาได้แก่หมู่บ้านในภาคเหนือ ภาคใต้และภาคกลาง ตามลำดับ และหมู่บ้านไฟเหลืองมีลักษณะการกระจายของความถี่ของการพัฒนาของหมู่บ้าน เป็น ภาคกลาง ภาคเหนือ ภาคตะวันออกเฉียงเหนือและ ภาคใต้ ตามลำดับ

สำหรับการสร้างตัวแบบเพื่อการจำแนกกลุ่มนั้น ทำได้ในทำนองเดียวกันกับตัวแบบที่ 1 โดยใช้เทคนิคการวิเคราะห์การจำแนกกลุ่ม (Discriminant Analysis) ซึ่งต้องตรวจสอบเงื่อนไขในการสร้างตัวแบบ รวมทั้งตรวจสอบความผิดปกติและการแจกแจงปกติของตัวแปรแต่ละตัวของข้อมูลในแต่ละกลุ่ม พร้อมกันนั้นได้ตรวจสอบความผิดปกติร่วมของตัวแปรของข้อมูลในแต่ละกลุ่ม โดยใช้สถิติ Mahalanobis distance ซึ่งจะใช้ค่านัยสำคัญที่ .001 ของการแจกแจงแบบไคกำลัง ที่ระดับองศาของความอิสระเท่ากับจำนวนตัวแปรที่ศึกษา ผลการศึกษาพบว่าในกลุ่มไฟสีเขียว เหลือข้อมูลอยู่จำนวน 10,010 หมู่บ้าน กลุ่มไฟสีเหลือง และกลุ่มไฟสีแดง เหลือข้อมูลอยู่จำนวน 7,893 หมู่บ้าน และ 7,560 หมู่บ้าน ตามลำดับ ผลการตรวจสอบการแจกแจงปกติและความผิดปกติของตัวแปรแต่ละตัวในแต่ละกลุ่มปรากฏในรูปที่ 2.1- 2.30 ในภาคผนวก ดังนั้นจึงเหลือข้อมูลที่จะใช้สำหรับการสร้างตัวแบบ Fisher Discriminant Function รวมทั้งสิ้น 25,463 หมู่บ้าน

สำหรับการตรวจสอบเงื่อนไขการเท่ากันของความผันแปรของตัวแปรในแต่ละกลุ่มพิจารณาได้จากรูปภาพที่ 2.31-2.33 ในภาคผนวก ซึ่งพบว่าการกระจายของคะแนนที่ได้จากตัวแบบของฟังก์ชันการจำแนกกลุ่มที่ได้จากกลุ่มไฟเขียว ไฟเหลือง และไฟแดง มีการกระจายที่พออนุโลมว่ามีความไม่แตกต่างกันมากนัก และนำไปสู่การสรุปเงื่อนไขของการเท่ากันของความผันแปรของตัวแปรในแต่ละกลุ่ม

ในการตรวจสอบ Multicollinearity ของตัวแปรอิสระ V24,V25,V26,V28, และ V29 โดยศึกษาจากสถิติสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปร ปรากฏในตารางที่ 2.3 ในภาคผนวก พบว่าค่าสัมประสิทธิ์สหสัมพันธ์สูงสุดได้แก่ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปร V24 กับ V28 คือ .694 ซึ่งเป็นค่าสัมประสิทธิ์สหสัมพันธ์ที่ไม่สูงมากนัก จึงอาจเชื่อได้ว่าจะไม่เกิด Multicollinearity ระหว่างตัวแปร

ดังนั้นโดยสรุปในการสร้างตัวแบบจะใช้ข้อมูลจากตัวแปรทั้งสิ้น 5 ตัวแปร ได้แก่ V24,V25, V26,V28, และ V29 จำนวน 25,463 หมู่บ้าน

จากการประมวลผลด้วยโปรแกรมสำเร็จรูปทางสถิติโดยใช้เทคนิคการจำแนกกลุ่ม Discriminant Analysis ได้ผลลัพธ์สุดท้ายดังตารางที่ 2.4-2.7 ในภาคผนวก ซึ่งพบว่า ค่าเฉลี่ยของตัวแปรอิสระแต่ละตัวมีความแตกต่างกันเมื่ออยู่ต่างกลุ่มกัน สัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรอิสระคู่ใด ๆ มีค่าค่อนข้างต่ำ จึงทำให้เชื่อได้ว่าจะไม่เกิดปัญหาที่เรียกว่า Multicollinearity ในตัวแบบ และจากค่าสัมประสิทธิ์ของตัวแปรในสมการจำแนกกลุ่ม ทำให้ได้ตัวแบบของการจำแนกกลุ่มเป็น

1. ตัวแบบการแบ่งกลุ่มของกลุ่มไฟเขียว คือ

$$D = -4.428 + 52.581 (V24) + 0.00000776 (V25) \\ + 0.0000133 (V26) + 0.000000706 (V28) + 0.905(V29)$$

2. ตัวแบบสำหรับการแบ่งกลุ่มของกลุ่มไฟเหลือง คือ

$$D = -18.718 + 90.517(V24) + 0.00000829(V25) \\ + 0.0000154(V26) + 0.0000256(V28) + 5.44(V29)$$

3. ตัวแบบสำหรับการแบ่งกลุ่มของกลุ่มไฟแดง คือ

$$D = -13.641 + 61.245(V24) + 0.0000164(V25) \\ + 0.0000491(V26) + 0.00000204(V18) + 1.595(V29)$$

ตารางที่ 2.7 ในภาคผนวก เป็นผลจากการใช้ตัวแบบที่ 2 พยากรณ์ข้อมูลรายหมู่บ้านตามความเสี่ยงของการพัฒนาเพื่อจัดเป็นหมู่บ้านไฟเขียว ไฟเหลือง หรือ ไฟแดง โดยวิธีการพยากรณ์ข้อมูลทั้งหมด(Original) และวิธีการพยากรณ์ข้อมูลบางส่วน(Cross-Validated) ผลปรากฏว่าทั้งสองวิธีให้ผลการพยากรณ์หมู่บ้านได้ถูกต้องประมาณ 97.6% เท่ากัน ซึ่งเป็นผลการพยากรณ์ที่อยู่ในระดับที่น่าพอใจ

ตัวแบบที่ 3 : ตัวแบบศักยภาพสำหรับการพัฒนา

จากรายงานการศึกษาต้นแบบกลไกการบูรณาการพัฒนาสังคม กำหนดตัวแปรที่ใช้พิจารณาศักยภาพสำหรับการพัฒนาไว้ทั้งหมด 15 ตัวแปร ได้แก่

1. จำนวนปีการศึกษาเฉลี่ยของประชากรที่ได้รับการศึกษา (V31)
2. จำนวนครัวเรือนที่มีอินเทอร์เน็ต (V32) *
3. จำนวนผู้สูงอายุที่มีงานทำ (V33)
4. รายได้จากการรับจ้างของครัวเรือนต่อปี (V34) *
5. จำนวนแรงงานที่มีการศึกษาเฉลี่ยมากกว่า 9 ปี (V35)
6. จำนวนผู้ทำกิจกรรมทางศาสนา (V36)
7. จำนวนสมาชิกในชมรมออกกำลังกาย (V37) *
8. จำนวนคนในครัวเรือนที่ไม่สูบบุหรี่ (V38)
9. ขนาดของครัวเรือนเฉลี่ย (V39)
10. ร้อยละของครัวเรือนที่จัดบ้านถูกลักษณะ มีการป้องกันอุบัติเหตุวิธี (V40) *
11. ร้อยละของคนในครัวเรือนได้รับรู้ข่าวสารที่เป็นประโยชน์ (V41) *
12. สัดส่วนพื้นที่หมู่บ้านต่อพื้นที่ทำการเกษตร พื้นที่ป่าสงวนและพื้นที่ปลูกป่า (V42) *
13. ร้อยละของครัวเรือนที่มีคนร่วมแสดงความคิดเห็นสาธารณะ (V43)*
14. ร้อยละของประชากรที่มีอายุมากกว่า 60 ปีขึ้นไป (V44)*
15. จำนวนประชากรทั้งหมด (V45)*

(* ตัวแปรที่เปลี่ยนไปจากตัวแปรเดิม)

ในทำนองเดียวกันกับตัวแบบที่ 1 และตัวแบบที่ 2 คณะที่ปรึกษาได้ทำการตรวจสอบความสมบูรณ์ของข้อมูลเบื้องต้นพบว่า มีจำนวนข้อมูล 4,194 หมู่บ้าน ที่มีค่าของตัวแปรสูญหาย และมีข้อมูลที่ไม่สมเหตุผลอีก 21 หมู่บ้าน จึงทำให้เหลือข้อมูล 64,979 หมู่บ้าน และจาก การพิจารณาความผิดปกติและการกระจายของข้อมูลรายตัวแปรพบว่า

- 1) ตัวแปร V33 มีค่าเท่ากันทั้งหมด และตัวแปร V37 ไม่สามารถหาข้อมูลรายหมู่บ้านได้ ดังนั้นตัวแปรดังกล่าวจึงไม่สามารถนำมาใช้เป็นตัวแปรที่จำแนกความแตกต่างของตัวแปรอื่นได้ จึงทำให้ต้องตัดตัวแปรเหล่านี้ทิ้งไป
- 2) ข้อมูลตัวแปร V32 มีค่าเท่ากับ 0 จำนวน 49,061 หมู่บ้าน ซึ่งประมาณ 75% ของข้อมูลทั้งหมด ถึงแม้ว่าจะเป็นจำนวนค่อนข้างมาก แต่ยังคงใช้ข้อมูลนี้ต่อไป
- 3) สำหรับข้อมูลของตัวแปร V31,V32,V34,V35, V36,V38,V39,V40, V41,V42,V43,V44 และ V45 พบว่าในแต่ละตัวแปรมีรายการข้อมูลที่ผิดปกติอยู่มีจำนวนแตกต่างกันไป ไม่มีจำนวนมากเหมือนตัวแปรในข้อ 2.)

จึงทำให้ต้องพิจารณาการตัดข้อมูลที่ผิดปกติบางหมู่บ้านออก เพื่อให้ได้ข้อมูลของทุกตัวแปรที่เหลือ มีค่าผิดปกติเหลือน้อยที่สุดหรือไม่เลย จากการพิจารณารายตัวแปรไปพร้อมๆกันทุกตัวแปร พบว่าสามารถกำจัดค่าผิดปกติของตัวแปรดังกล่าวได้เกือบทั้งหมด แต่จะเหลือข้อมูลที่ใช้พิจารณาสร้างตัวแบบทั้งสิ้น 3,004 หมู่บ้านเท่านั้น ซึ่งในจำนวนนี้ไม่สามารถกำจัดค่าผิดปกติของตัวแปร V32 ได้ จึงมีความจำเป็นต้องตัดตัวแปร V32 ทิ้ง

4) โดยสรุปมีตัวแปรที่ใช้พิจารณาจัดกลุ่มในตัวแบบที่ 3 จำนวน 12 ตัวแปร และมีข้อมูลจำนวนทั้งสิ้น 3,004 หมู่บ้าน

การพัฒนาตัวแบบเพื่อการจำแนกกลุ่ม

ในทำนองเดียวกันกับการพัฒนาตัวแบบที่ 1 และตัวแบบที่ 2 การพิจารณาตัวแบบที่ 3 มีความจำเป็นต้องมีการวิเคราะห์เพื่อการจัดกลุ่มข้อมูลเสียก่อน ซึ่งกระทรวงการพัฒนาสังคมและความมั่นคงของมนุษย์ได้แบ่งหมู่บ้านตามศักยภาพสำหรับการพัฒนาออกเป็น 3 กลุ่ม ได้แก่

กลุ่มหมู่บ้านที่มีศักยภาพสำหรับการพัฒนาสูง กำหนดให้เป็นกลุ่มไฟเขียว

กลุ่มหมู่บ้านที่มีศักยภาพสำหรับการพัฒนาปานกลาง กำหนดให้เป็นกลุ่มไฟเหลือง

กลุ่มหมู่บ้านที่มีศักยภาพสำหรับการพัฒนาต่ำ กำหนดให้เป็นกลุ่มไฟแดง

และจากข้อมูล 12 ตัวแปร จำนวน 3,004 หมู่บ้าน คณะผู้ศึกษาทำการวิเคราะห์เพื่อการจัดกลุ่มข้อมูลหมู่บ้านด้วยเทคนิค Cluster Analysis ด้วยโปรแกรมสำเร็จรูปทางสถิติได้ผลลัพธ์การจัดกลุ่มในรอบสุดท้ายของการจัดกลุ่มดังตารางที่ 3.1 และ 3.2 ในภาคผนวก

จากตารางที่ 3.1 และ 3.2 และความเห็นของผู้เชี่ยวชาญทางด้านสังคมศาสตร์และความมั่นคงของมนุษย์ จัดกลุ่มหมู่บ้านได้เป็น

กลุ่มที่ 1	กลุ่มไฟเขียว	มีหมู่บ้านที่จัดอยู่ในกลุ่มนี้จำนวนทั้งหมด	1,092	หมู่บ้าน
กลุ่มที่ 2	กลุ่มไฟเหลือง	มีหมู่บ้านที่จัดอยู่ในกลุ่มนี้จำนวนทั้งหมด	890	หมู่บ้าน
กลุ่มที่ 3	กลุ่มไฟแดง	มีหมู่บ้านที่จัดอยู่ในกลุ่มนี้จำนวนทั้งหมด	1,022	หมู่บ้าน

ตารางที่ 3 จำนวนหมู่บ้านจำแนกตามประเภทของกลุ่มและภาคต่างๆ ของตัวแบบที่ 3

กลุ่ม	ภาค				รวม (%)
	เหนือ (%)	กลาง (%)	ตะวันออกเฉียงเหนือ (%)	ใต้ (%)	
ไฟเขียว	302 (10.1)	103 (3.4)	540 (18.0)	147 (4.9)	1092 (36.4)
ไฟเหลือง	393 (13.1)	175 (5.8)	137 (4.6)	185 (6.2)	890 (29.6)
ไฟแดง	171 (5.7)	57 (1.9)	764 (25.4)	30 (1.0)	8342 (23.1)
รวม	866 (28.8)	335 (11.2)	1441 (48.0)	362 (12.1)	3004 (100.0)

จากตารางที่ 3 ข้างต้น พบว่าหมู่บ้านที่จัดอยู่ในกลุ่มไฟแดง อยู่ในภาคตะวันออกเฉียงเหนือสูงสุด (25.4%) รองลงมาได้แก่หมู่บ้านในภาคเหนือ ภาคกลางและภาคใต้ ตามลำดับ สำหรับหมู่บ้านที่จัดเป็นกลุ่มไฟเขียวอยู่ในภาคตะวันออกเฉียงเหนือสูงสุด (18.0%) รองลงมาได้แก่หมู่บ้านในภาคเหนือ ภาคใต้และภาคกลาง ตามลำดับ สำหรับหมู่บ้านที่จัดเป็นกลุ่มไฟเหลืองมีลักษณะการกระจายของศักยภาพสำหรับการพัฒนาของหมู่บ้าน เป็นภาคเหนือ ภาคใต้ ภาคกลาง และ ภาคตะวันออกเฉียงเหนือ ตามลำดับ

สำหรับการสร้างตัวแบบเพื่อการจำแนกกลุ่มของกลุ่มตัวแปรศักยภาพสำหรับการพัฒนา พัฒนาในตนเอง เดียวกันกับตัวแบบที่ 1 และตัวแบบที่ 2 โดยใช้เทคนิค Discriminant Analysis ซึ่งต้องทำการตรวจสอบเงื่อนไขของตัวแบบ ทั้งความผิดปกติและการแจกแจงปกติของตัวแปรแต่ละตัวของข้อมูลในแต่ละกลุ่ม พร้อมกันนั้นได้ตรวจสอบความผิดปกติร่วมของตัวแปรของข้อมูลในแต่ละกลุ่ม โดยใช้สถิติ Mahalanobis distance ที่ระดับนัยสำคัญ .001 ผลปรากฏว่าเหลือข้อมูลในกลุ่มไฟสีเขียวอยู่จำนวน 877 หมู่บ้าน ในกลุ่มสีเหลือง และกลุ่มสีแดงเหลืออยู่ 862 หมู่บ้าน และ 774 หมู่บ้าน ตามลำดับ และเนื่องจากไม่สามารถกำจัดค่าผิดปกติของตัวแปร V41 ออกได้ จึงจำเป็นต้องตัดตัวแปร V41 ออกจากการพิจารณา ผลการตรวจสอบการแจกแจงปกติและความผิดปกติของตัวแปรแต่ละตัวในแต่ละกลุ่มปรากฏในรูปที่ 3.1- 3.66 ในภาคผนวก โดยสรุปจะเหลือข้อมูลที่จะใช้สำหรับการสร้างตัวแบบ Fisher Discriminant Function รวมทั้งสิ้น 2,513 หมู่บ้าน 11 ตัวแปร คือ V31,V34,V35,V36,V38,V39,V40,V42,V43,V44 และ V45

สำหรับการตรวจสอบเงื่อนไขการเท่ากันของความแปรของตัวแปรในแต่ละกลุ่มพิจารณาได้จากรูปภาพที่ 3.67-3.69 ในภาคผนวก ซึ่งพบว่าการกระจายของคะแนนที่ได้จากตัวแบบของฟังก์ชันการจำแนกกลุ่มที่ได้จาก

กลุ่มไฟเขียว ไฟเหลือง และไฟแดง มีการกระจายที่พออนุโลมว่ามีความไม่แตกต่างกันมากนัก จึงนำไปสู่การสรุปเงื่อนไขของการเท่ากันของความผันแปรของตัวแปรในแต่ละกลุ่ม

ในการตรวจสอบ Multicollinearity ของตัวแปรอิสระ V31,V34,V35,V36,V38,V39,V40,V42,V43,V44 และ V45 โดยศึกษาจากความสัมพันธ์ระหว่างตัวแปรจากสถิติสัมประสิทธิ์สหสัมพันธ์ ปรากฏในตารางที่ 3.3 ในภาคผนวก พบว่า V38 มีความสัมพันธ์กับ V45 สูงสุด คือ .978 ตัวแปร V36 มีความสัมพันธ์กับ V45 ในระดับ .951 และ ตัวแปร V36 มีความสัมพันธ์กับ V38 ในระดับ .943 ซึ่งมีความสัมพันธ์กันในระดับที่สูงมาก เมื่อพิจารณาความหมายของตัวแปร พบว่าศักยภาพสำหรับการพัฒนาน่าจะเกี่ยวข้องอยู่กับจำนวนประชากร (V45) มากกว่าจำนวนผู้ทำกิจกรรมทางศาสนา(V36)และจำนวนคนในครอบครัวที่ไม่สูบบุหรี่(V38) จึงตัดตัวแปร V36 และ V38 ออกจาก การพิจารณาสร้างตัวแบบ ทำให้เหลือตัวแปรที่จะใช้ในการสร้างตัวแบบทั้งสิ้น 9 ตัวแปร ได้แก่ V31,V34,V35, V39,V40,V42,V43,V44 และ V45 โดยใช้ข้อมูลรายหมู่บ้านทั้งสิ้น 2,513 หมู่บ้าน

จากการประมวลผลด้วยโปรแกรมสำเร็จรูปทางสถิติโดยใช้เทคนิคการจำแนกกลุ่ม Discriminant Analysis ได้ผลลัพธ์สุดท้ายดังตารางที่ 3.4-3.7 ในภาคผนวก ซึ่งพบว่า ค่าเฉลี่ยของตัวแปรอิสระแต่ละตัวมีความแตกต่างกันเมื่ออยู่ต่างกลุ่มกัน ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรอิสระคู่ใด ๆ มีค่าค่อนข้างต่ำ จึงทำให้เชื่อได้ว่าจะไม่เกิดปัญหา Multicollinearity ในตัวแบบ และจากตารางที่ 3.6 ค่าสัมประสิทธิ์ของตัวแปรในสมการจำแนกกลุ่ม ซึ่งพบว่าในขั้นตอนการเลือกตัวแปรเพื่อการสร้างตัวแบบ(Stepwise method) ตัวแปร V42 ไม่ถูกเลือกในการสร้างตัวแบบ ดังนั้นตัวแบบการจำแนกกลุ่มจึงเป็น

1. ตัวแบบการแบ่งกลุ่มของกลุ่มไฟเขียว คือ

$$D = -11071.5 + 109.336(V31) - 0.000026(V34) + 0.493 (V35) \\ + 48.371 (V39) + 198.308 (V40) + 13.954(V43) - 2.217 (V44) \\ - 0.192(V45)$$

2. ตัวแบบสำหรับการแบ่งกลุ่มของกลุ่มไฟเหลือง คือ

$$D = -10949.5 + 110.357(V31) - 0.000021(V34) + 0.526 (V35) \\ + 46.651 (V39) + 197.457 (V40) + 13.613(V43) - 1.977 (V44) \\ - 0.220(V45)$$

3. ตัวแบบสำหรับการแบ่งกลุ่มของกลุ่มไฟแดง คือ

$$D = -10930.5 + 107.936(V31) - 0.000034(V34) + 0.466 (V35) \\ + 49.138 (V39) + 197.014 (V40) + 14.083(V43) - 2.304 (V44) \\ - 0.218(V45)$$

ตารางที่ 3.7 ในภาคผนวก เป็นผลจากการใช้ตัวแบบที่สร้างขึ้นไปพยากรณ์ข้อมูลรายหมู่บ้านตามศักยภาพ
สำหรับการพัฒนาเพื่อจัดเป็นหมู่บ้านไฟฟ้าชีว ไฟเหลือง หรือ ไฟแดง ทั้งโดยวิธีการพยากรณ์ข้อมูลทั้งหมด
(Original) และการพยากรณ์ข้อมูลบางส่วน(Cross-Validated) ผลปรากฏว่าทั้งสองวิธีให้ผลการพยากรณ์หมู่บ้าน
ได้ถูกต้องใกล้เคียงกันคือประมาณ 95% เท่ากัน ซึ่งแสดงว่าผลการพยากรณ์ของตัวแบบที่สร้างขึ้นอยู่ในระดับที่น่า
พอใจ
