

บทที่ 3

วิธีดำเนินการวิจัย

การวิจัยครั้งนี้เป็นการศึกษาวิธีการในการคัดเลือกตัวแบบการถดถอยเชิงเส้นพหุคูณ โดยวิธี DGA และเปรียบเทียบวิธีดังกล่าวกับวิธี SR และวิธี GA โดยใช้เกณฑ์ AIC เป็นเกณฑ์ในการพิจารณาเลือกตัวแบบ และเกณฑ์ MSE ซึ่งประกอบด้วยค่า MSE ที่ต่ำที่สุดและค่า AMSE เป็นเกณฑ์ในการเปรียบเทียบประสิทธิภาพของทั้ง 3 วิธี สำหรับข้อมูลที่ใช้ในการวิจัยได้จากการจำลองสถานการณ์โดยโปรแกรม R Version 3.0.1 และ R Commander Version 1.9-6 ผู้วิจัยได้กำหนดการดำเนินงานต่างๆ ดังนี้

3.1 แผนการดำเนินการวิจัย

3.1.1 ศึกษาภายใต้ตัวแบบการถดถอยเชิงเส้นพหุคูณที่อยู่ในรูปเชิงเส้นของพารามิเตอร์และตัวแปรอิสระโดยมีรูปแบบทั่วไปคือ

$$y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} + \varepsilon_i$$

โดยที่ y_i แทน ตัวแปรตาม เมื่อ $i = 1, \dots, n$ และ n แทนขนาดตัวอย่าง

$X_{i1}, X_{i2}, \dots, X_{ik}$ แทน ตัวแปรอิสระ เมื่อ k แทนจำนวนตัวแปรอิสระ

$\beta_0, \beta_1, \dots, \beta_k$ แทน สัมประสิทธิ์การถดถอย

ε_i แทน ความคลาดเคลื่อนสุ่ม

3.1.2 ทำการจำลองข้อมูล เพื่อสร้างตัวแบบถดถอยโดยกำหนดตัวแปรอิสระมีการแจกแจง

แบบเอกรูป มีรูปแบบคือ $X_{ij} \sim U(a, b)$ กำหนดความคลาดเคลื่อนมีการแจกแจงแบบปกติ มีรูปแบบคือ $\varepsilon_i \sim N(0, \sigma^2)$ และกำหนดขนาดประชากร $N=10,000$

3.1.3 ทำการศึกษาภายใต้สถานการณ์ต่างๆ ดังนี้

1) การเกิดปัญหาสหสัมพันธ์เชิงเส้นพหุและไม่เกิดปัญหาสหสัมพันธ์เชิงเส้นพหุ

2) ขนาดตัวอย่างเท่ากับ 50, 100 และ 500

3) กำหนดความคลาดเคลื่อนโดย $\varepsilon_i \sim N(0, 50^2)$, $\varepsilon_i \sim N(0, 80^2)$,
 $\varepsilon_i \sim N(0, 150^2)$ และ $\varepsilon_i \sim N(0, 250^2)$

4) กำหนดจำนวนตัวแปรอิสระเท่ากับ 10 และ 20 ตัวแปร

3.1.4 กำหนดค่าสัมประสิทธิ์การถดถอยโดยพิจารณาจากช่วงของตัวแปรอิสระแต่ละตัวที่ทำการกำหนดขึ้นและจากผลงานวิจัยของ กานต์ณัฐ ณ บางช้าง (5) ที่ทำการศึกษาเกี่ยวกับการคัดเลือกแปรในตัวแบบการถดถอยเชิงเส้นซึ่งมีลักษณะที่คล้ายกัน ดังนั้นจึงทำการกำหนดค่าสัมประสิทธิ์การถดถอย เมื่อตัวแปรอิสระมี 10 ตัวแปร และ 20 ตัวแปร ตามรูปแบบที่ได้กำหนดขึ้น

3.1.5 กำหนดตัวแปรตามจากตัวแบบการถดถอยที่ตัวแปรอิสระมีจำนวน 10 ตัวแปรและ 20 ตัวแปร ตามรูปแบบที่ได้กำหนดขึ้น

3.1.6 สำหรับสถานการณ์ที่เกิดปัญหาสหสัมพันธ์เชิงเส้นพหุ กำหนดให้ตัวแปรอิสระ

$$X_{i3} = X_{i1} + X_{i2}$$

3.1.7 วิธีการคัดเลือกตัวแปรที่ใช้ ได้แก่ วิธี DGA วิธี SR และวิธี GA

3.1.8 ฟังก์ชันเป้าหมายของวิธี DGA ได้แก่ ฟังก์ชัน AIC และ MSE ส่วนฟังก์ชันเป้าหมายของวิธี GA ได้แก่ ฟังก์ชัน AIC

3.1.9 ทำการคัดเลือกตัวแบบโดยใช้เกณฑ์ AIC และทำการเปรียบเทียบวิธี DGA วิธี SR และวิธี GA โดยใช้เกณฑ์ MSE ซึ่งประกอบด้วยค่า MSE ที่ต่ำที่สุดและค่า AMSE

3.1.10 กำหนดจำนวนกระทำซ้ำ 100 รอบ และ จำนวนรุ่นประชากรของวิธี GA และวิธี DGA เท่ากับ 2,000 รุ่น ในแต่ละสถานการณ์ จากสมการต่อไปนี้

$$R \geq \left(\frac{t_{\frac{\alpha}{2}, R_0 - 1} S_0}{\varepsilon} \right)^2$$

โดย R แทน จำนวนรอบในการประมวลผล

R_0 แทน จำนวนรอบในการประมวลผลเริ่มต้น

α แทน ระดับนัยสำคัญ

$t_{\frac{\alpha}{2}, R_0 - 1}$ แทน ค่าการแจกแจงที

ε แทน ค่าความคลาดเคลื่อนที่ยอมรับได้

S_0 แทน ค่าเบี่ยงเบนมาตรฐาน

3.1.11 ทำการเปรียบเทียบการคัดเลือกตัวแบบการถดถอยทั้ง 3 วิธี และสรุปผลที่ได้

3.2 ขั้นตอนการดำเนินการวิจัย

การดำเนินการวิจัยมีขั้นตอนที่สำคัญตามลำดับดังนี้

3.2.1 การจำลองข้อมูล

1) สถานการณ์ที่ไม่เกิดปัญหาสหสัมพันธ์เชิงเส้นพหุ

1.1) กำหนดขนาดประชากร $N = 10,000$

1.2) ตัวแปรอิสระจำนวน 10 ตัวแปร มีการแจกแจงแบบเอกรูป ดังนี้

$$X_{i1} \sim U(25, 200), X_{i2} \sim U(50, 180), X_{i3} \sim U(15, 250), \\ X_{i4} \sim U(87, 550), X_{i5} \sim U(1, 70), X_{i6} \sim U(35, 90), X_{i7} \sim U(3, 17), \\ X_{i8} \sim U(8, 22), X_{i9} \sim U(15, 80) \text{ และ } X_{i10} \sim U(20, 150)$$

1.3) ตัวแปรอิสระจำนวน 20 ตัวแปร มีการแจกแจงแบบเอกรูป ดังนี้

$$X_{i1} \sim U(25, 200), X_{i2} \sim U(50, 180), X_{i3} \sim U(15, 250), \\ X_{i4} \sim U(87, 550), X_{i5} \sim U(1, 70), X_{i6} \sim U(35, 90), X_{i7} \sim U(3, 17), \\ X_{i8} \sim U(8, 22), X_{i9} \sim U(15, 80), X_{i10} \sim U(20, 150), \\ X_{i11} \sim U(1200, 1550), X_{i12} \sim U(800, 900), X_{i13} \sim U(890, 1000), \\ X_{i14} \sim U(500, 900), X_{i15} \sim U(1300, 1700), X_{i16} \sim U(3500, 5000), \\ X_{i17} \sim U(3200, 17000), X_{i18} \sim U(1900, 2900), \\ X_{i19} \sim U(15000, 25000) \text{ และ } X_{i20} \sim U(200000, 500000)$$

1.4) ความคลาดเคลื่อนมีการแจกแจงปกติ โดยกำหนดให้

$$\varepsilon_i \sim N(0, 50^2), \varepsilon_i \sim N(0, 80^2), \varepsilon_i \sim N(0, 150^2) \text{ และ } \varepsilon_i \sim N(0, 250^2)$$

1.5) ค่าสัมประสิทธิ์การถดถอยของตัวแปรอิสระ 10 ตัวแปร ดังนี้

$$\beta_0 = 100, \beta_1 = -155, \beta_2 = 15, \beta_3 = 11.5, \beta_4 = 509, \beta_5 = 0, \beta_6 = 0, \\ \beta_7 = 4, \beta_8 = 7, \beta_9 = 0 \text{ และ } \beta_{10} = 0$$

1.6) ค่าสัมประสิทธิ์การถดถอยของตัวแปรอิสระ 20 ตัวแปร ดังนี้

$$\beta_0 = 100, \beta_1 = -155, \beta_2 = 15, \beta_3 = 11.5, \beta_4 = 509, \beta_5 = 0, \beta_6 = 0, \\ \beta_7 = 4, \beta_8 = 7, \beta_9 = 0, \beta_{10} = 0, \beta_{11} = -1550, \beta_{12} = 900, \beta_{13} = 1000, \\ \beta_{14} = 50, \beta_{15} = 0, \beta_{16} = 0, \beta_{17} = 40, \beta_{18} = 100, \beta_{19} = 0 \text{ และ } \beta_{20} = 0$$

1.7) สุ่มขนาดตัวอย่างจำนวน 50, 100 และ 500

1.8) สร้างตัวแปรตาม จากตัวแบบการถดถอยต่อไปนี้

เมื่อตัวแปรอิสระ 10 ตัวแปร ดังนี้

$$y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \dots + \beta_9 X_{i9} + \beta_{10} X_{i10} + \varepsilon_i$$

โดยที่ y_i แทน ตัวแปรตาม เมื่อ $i = 1, \dots, n$ และ n แทนขนาดตัวอย่าง

$X_{i1}, X_{i2}, \dots, X_{i10}$ แทน ตัวแปรอิสระ

$\beta_0, \beta_1, \dots, \beta_{10}$ แทน สัมประสิทธิ์การถดถอย

ε_i แทน ความคลาดเคลื่อนสุ่ม

เมื่อตัวแปรอิสระ 20 ตัวแปร ดังนี้

$$y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \dots + \beta_{19} X_{i19} + \beta_{20} X_{i20} + \varepsilon_i$$

โดยที่ $X_{i1}, X_{i2}, \dots, X_{i20}$ แทน ตัวแปรอิสระ

$\beta_0, \beta_1, \dots, \beta_{20}$ แทน สัมประสิทธิ์การถดถอย

2) สถานการณ์ที่เกิดปัญหาสหสัมพันธ์เชิงเส้นพหุ

2.1) กำหนดขนาดประชากร $N = 10,000$

2.2) ตัวแปรอิสระจำนวน 10 ตัวแปร มีการแจกแจงแบบเอกรูป ดังนี้

$$X_{i1} \sim U(25, 200), X_{i2} \sim U(50, 180), X_{i3} \sim U(15, 250),$$

$$X_{i4} \sim U(87, 550), X_{i5} \sim U(1, 70), X_{i6} \sim U(35, 90), X_{i7} \sim U(3, 17),$$

$$X_{i8} \sim U(8, 22), X_{i9} \sim U(15, 80) \text{ และ } X_{i10} \sim U(20, 150)$$

2.3) ตัวแปรอิสระจำนวน 20 ตัวแปร มีการแจกแจงแบบเอกรูป ดังนี้

$$\begin{aligned} X_{i1} &\square U(25,200), X_{i2} \square U(50,180), X_{i3} \square U(15,250), \\ X_{i4} &\square U(87,550), X_{i5} \square U(1,70), X_{i6} \square U(35,90), X_{i7} \square U(3,17), \\ X_{i8} &\square U(8,22), X_{i9} \square U(15,80), X_{i10} \square U(20,150), \\ X_{i11} &\square U(1200,1550), X_{i12} \square U(800,900), X_{i13} \square U(890,1000), \\ X_{i14} &\square U(500,900), X_{i15} \square U(1300,1700), X_{i16} \square U(3500,5000), \\ X_{i17} &\square U(3200,17000), X_{i18} \square U(1900,2900), \\ X_{i19} &\square U(15000,25000) \text{ และ } X_{i20} \square U(200000,500000) \end{aligned}$$

2.4) ความคลาดเคลื่อนมีการแจกแจงปกติ โดยกำหนดให้

$$\varepsilon_i \square N(0,50^2), \varepsilon_i \square N(0,80^2), \varepsilon_i \square N(0,150^2) \text{ และ } \varepsilon_i \square N(0,250^2)$$

2.5) ค่าสัมประสิทธิ์การถดถอยของตัวแปรอิสระ 10 ตัวแปร ดังนี้

$$\beta_0 = 100, \beta_1 = -155, \beta_2 = 15, \beta_3 = 11.5, \beta_4 = 509, \beta_5 = 0, \beta_6 = 0, \\ \beta_7 = 4, \beta_8 = 7, \beta_9 = 0 \text{ และ } \beta_{10} = 0$$

2.6) ค่าสัมประสิทธิ์การถดถอยของตัวแปรอิสระ 20 ตัวแปร ดังนี้

$$\begin{aligned} \beta_0 &= 100, \beta_1 = -155, \beta_2 = 15, \beta_3 = 11.5, \beta_4 = 509, \beta_5 = 0, \beta_6 = 0, \\ \beta_7 &= 4, \beta_8 = 7, \beta_9 = 0, \beta_{10} = 0, \beta_{11} = -1550, \beta_{12} = 900, \beta_{13} = 1000, \\ \beta_{14} &= 50, \beta_{15} = 0, \beta_{16} = 0, \beta_{17} = 40, \beta_{18} = 100, \beta_{19} = 0 \text{ และ } \beta_{20} = 0 \end{aligned}$$

2.7) สุ่มขนาดตัวอย่างจำนวน 50, 100 และ 500

2.8) สร้างตัวแปรตามจากตัวแปรอิสระ 10 ตัวแปรและตัวแปรอิสระ 20 ตัวแปร จาก
ตัวแบบการถดถอยที่กำหนดไว้

2.9) กำหนดความสัมพันธ์ระหว่างตัวแปรอิสระกำหนดให้ X_{i1}, X_{i3} และ X_{i2}, X_{i3}

$$\text{มีความสัมพันธ์กันโดยการสร้างตัวแปรอิสระในรูปแบบ } X_{i3} = X_{i1} + X_{i2}$$

3.2.2 การคัดเลือกตัวแบบการถดถอย

1) ขั้นตอนการคัดเลือกตัวแบบโดยวิธี SR

ขั้นที่ 1 เริ่มจากตัวแบบที่ไม่มีตัวแปรอิสระตัวใดอยู่ในตัวแบบ

ขั้นที่ 2 ทำการพิจารณาเลือกตัวแปรอิสระตัวแรกเข้าในตัวแบบการถดถอย โดยพิจารณาจากค่า AIC ที่ต่ำที่สุด นั่นคือ การพิจารณาค่า AIC ที่ตัวแปรเหล่านั้น อยู่ในตัวแบบ หากตัวแปรอิสระตัวใดอยู่ในตัวแบบแล้วทำให้ได้ค่า AIC ที่ต่ำที่สุด ก็จะทำการเลือกตัวแปรนั้นเป็นตัวแรกของตัวแบบการถดถอย

ขั้นที่ 3 พิจารณาเลือกตัวแปรอิสระตัวถัดไปเข้าสู่ตัวแบบเมื่อมีตัวแปรอิสระ ก่อนหน้านี้ที่อยู่ในตัวแบบด้วย โดยพิจารณาค่า AIC ของตัวแบบที่มีตัวแปรที่ยัง เหลืออยู่แต่ละตัวในตัวแบบ แล้วให้ค่า AIC ต่ำที่สุด จะทำการเลือกตัวแบบนั้น เพื่อ เข้าสู่ขั้นตอนต่อไป

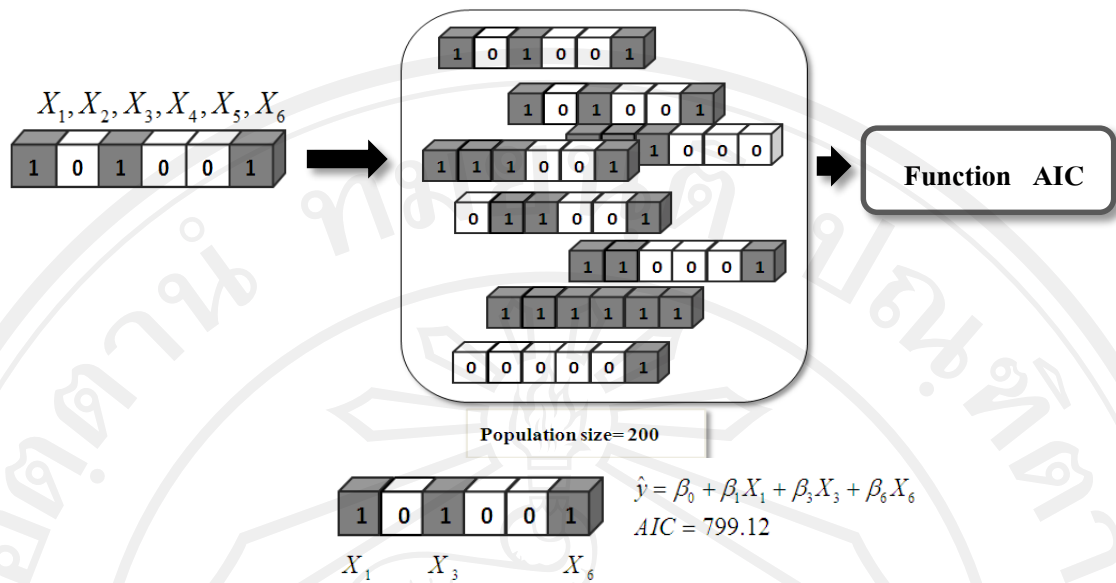
ขั้นที่ 4 จากตัวแบบการถดถอยที่ได้จะทำการพิจารณาค่า AIC ของตัวแบบ เมื่อทำการตัดตัวแปรอิสระก่อนหน้านี้ออก หากค่า AIC ของตัวแบบนั้นมีค่าที่ลดลง จะทำการหยุดกระบวนการทำงาน ซึ่งแสดงว่าตัวแบบที่ได้นี้ เป็นตัวแบบที่ เหมาะสมที่สุดโดยให้ค่า AIC ต่ำที่สุดแล้ว หากมีค่ามากกว่า จะเข้าสู่ขั้นตอนต่อไป

ขั้นที่ 5 กระทำซ้ำในขั้นที่ 3 และขั้นที่ 4 จนกว่าจะไม่มีตัวแปรอิสระตัวใด เข้าสู่ตัวแบบหรือถูกตัดออกจากตัวแบบอีก

ขั้นที่ 6 นำตัวแบบที่ได้มาทำการประมาณค่าสัมประสิทธิ์การถดถอยโดยวิธี OLS จากนั้นคำนวณค่า MSE ของตัวแบบที่ได้

2) ขั้นตอนการคัดเลือกตัวแบบโดยวิธี GA

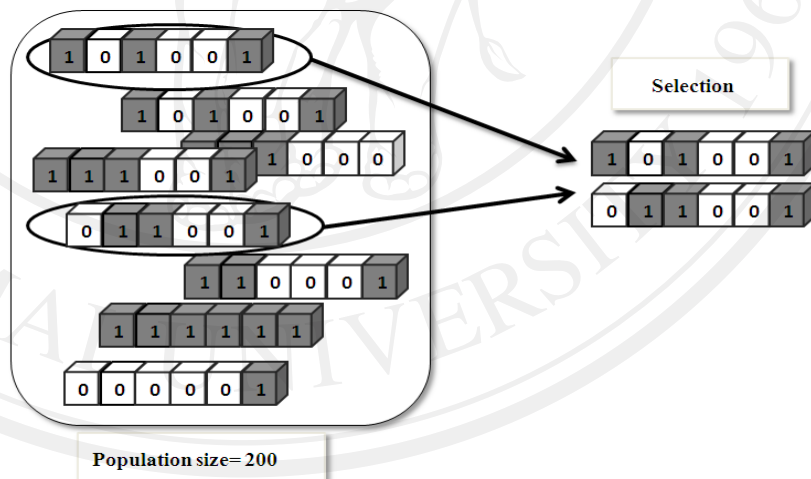
ขั้นที่ 1 สร้างประชากรขึ้นมาจากการสุ่มเพื่อเป็นประชากรเริ่มต้น ของ ประชากรรุ่นที่ 1 โดยอยู่ในลักษณะสายโครโมโซมประกอบด้วยยีนแบบไบนารี จำนวน 200 สายโครโมโซม โดยจำนวนยีนจะเท่ากับจำนวนตัวแปรอิสระ พร้อม ทั้งแปลงรหัสเพื่อคำนวณค่า AIC ของแต่ละโครโมโซม (28) ดังแสดงในรูป 3.1



รูป 3.1 แสดงรูปแบบสายโครโมโซม จำนวนโครโมโซมที่สร้างขึ้นและการแปลรหัส

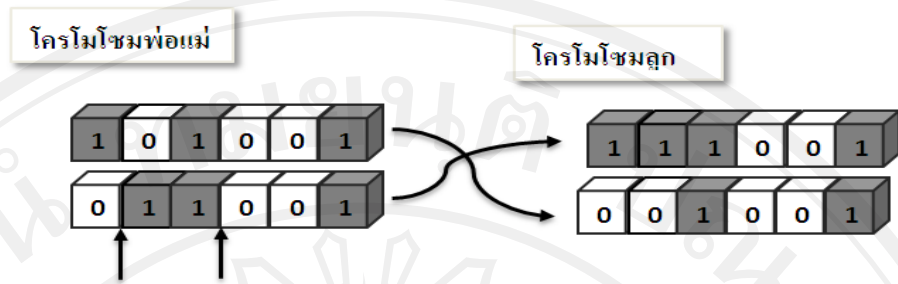
ขั้นที่ 2 ทำการคัดเลือกประชากรเพื่อเป็นโครโมโซมต้นแบบหรือเรียกว่าโครโมโซมพ่อแม่จากโครโมโซมที่มีค่า AIC ที่น้อยที่สุด 2 อันดับ ดังแสดงในรูป

3.2



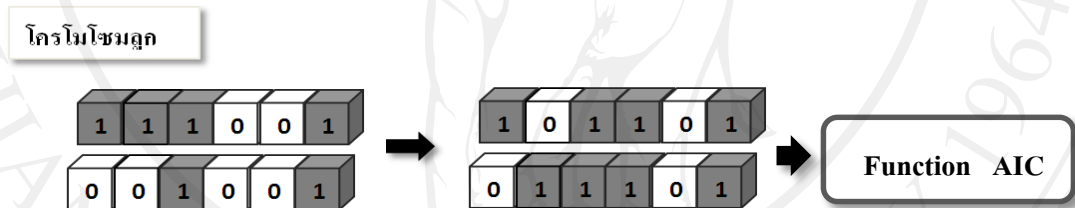
รูป 3.2 แสดงการคัดเลือกโครโมโซม

ขั้นที่ 3 กำหนดความน่าจะเป็นของการสลับสายพันธุ เท่ากับ 0.8 จากนั้นสุ่มตัวเลขในช่วง 0 ถึง 1 หากมีค่าน้อยกว่าหรือเท่ากับ 0.8 จะทำการสลับสายพันธุ ในที่นี้จะทำการ สลับสายพันธุ 2 จุด จากนั้นสุ่มเพื่อเลือกตำแหน่งที่จะสลับสายพันธุ พร้อมทั้งทำการสลับสายพันธุ ดังแสดงในรูป 3.3



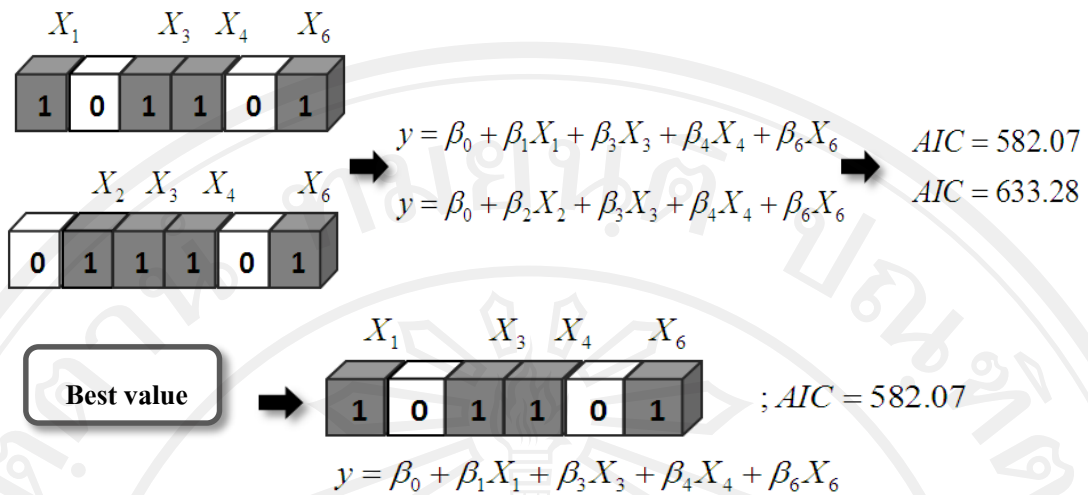
รูป 3.3 แสดงการสลับสายพันธุ์ของโครโมโซม

ขั้นที่ 4 กำหนดความน่าจะเป็นของการกลายพันธุ์ เท่ากับ 0.2 จากนั้นสุ่มตัวเลขในช่วง 0 ถึง 1 หากมีค่าน้อยกว่าหรือเท่ากับ 0.2 จะทำการกลายพันธุ์ ในกรณีที่มีการกลายพันธุ์ ตำแหน่งและจำนวนตำแหน่งที่จะทำการกลายพันธุ์ เป็นไปโดยสุ่ม จากนั้นจึงทำการกลายพันธุ์ พร้อมทั้งคำนวณค่า AIC ของโครโมโซมที่ได้ ดังแสดงในรูป 3.4



รูป 3.4 แสดงการกลายพันธุ์ของโครโมโซม

ขั้นที่ 5 ทำการแปลงรหัสจากโครโมโซมที่ได้มาเป็นตัวแปรที่เหมาะสมจะอยู่ในตัวแบบการถดถอยพร้อมทั้งคำนวณค่า AIC ทำการเลือกโครโมโซมที่ให้ค่า AIC น้อยที่สุด เพื่อนำมาเป็นตัวแบบที่เหมาะสมที่สุดสำหรับ ประชากรรุ่นที่ 1 ดังแสดงในรูป 3.5 โดยนำโครโมโซมพ่อแม่มาพิจารณาด้วย ดำเนินการทั้งสิ้นจำนวน 2,000 รุ่น แล้วนำโครโมโซมในรุ่นที่ 2,000 มาดำเนินการในขั้นต่อไป



รูป 3.5 แสดงการแปลงรหัสของโครโมโซมของประชากรรุ่นที่ 1 และเลือกโครโมโซม

ขั้นที่ 6 เมื่อได้โครโมโซมที่ให้ค่า AIC ต่ำที่สุดแล้ว ทำการประมาณค่าสัมประสิทธิ์การถดถอยโดยวิธี OLS จากนั้นคำนวณค่า MSE

3) ขั้นตอนการคัดเลือกตัวแบบโดยวิธี DGA

ประกอบด้วย 2 ส่วน ซึ่งแต่ละส่วนจะมีขั้นตอนย่อยโดยในส่วนที่ 1 ขั้นที่ 1 ถึง ขั้นที่ 5 มีกระบวนการทำงานคล้ายกับวิธี GA ดังนี้

ส่วนที่ 1

ขั้นที่ 1 สร้างประชากรขึ้นมาจากการสุ่มเพื่อเป็นประชากรเริ่มต้น ของประชากรรุ่นที่ 1 โดยอยู่ในลักษณะสายโครโมโซมประกอบด้วยยีนแบบไบนารี จำนวน 200 สายโครโมโซม โดยจำนวนยีนจะเท่ากับจำนวนตัวแปรอิสระ พร้อมทั้งแปลงรหัสเพื่อคำนวณค่า AIC ของแต่ละโครโมโซม ดังแสดงในรูป 3.1

ขั้นที่ 2 ทำการคัดเลือก ประชากรเพื่อเป็นโครโมโซมต้นแบบหรือเรียกว่าโครโมโซมพ่อแม่ จากโครโมโซมที่มีค่า AIC ที่น้อยที่สุด 2 อันดับ ดังแสดงในรูป 3.2

ขั้นที่ 3 กำหนดความน่าจะเป็นของการสลับสายพันธุ์ เท่ากับ 0.8 จากนั้นสุ่มตัวเลขในช่วง 0 ถึง 1 หากมีค่าน้อยกว่าหรือเท่ากับ 0.8 จะทำการสลับสายพันธุ์ ใน

ที่นี่จะทำการ สลับสายพันธุ์ 2 จุด จากนั้นสุ่มเพื่อเลือกตำแหน่งที่จะ สลับสายพันธุ์ พร้อมทั้งทำการสลับสายพันธุ์ ดังแสดงในรูป 3.3

ขั้นที่ 4 กำหนดความน่าจะเป็นของการกลายพันธุ์ เท่ากับ 0.2 จากนั้นสุ่มตัวเลขในช่วง 0 ถึง 1 หากมีค่าน้อยกว่าหรือเท่ากับ 0.2 จะทำการกลายพันธุ์ ในกรณีที่มีการกลายพันธุ์ ตำแหน่งและจำนวนตำแหน่งที่จะทำการกลายพันธุ์ เป็นไปโดยสุ่ม จากนั้นจึงทำการกลายพันธุ์ พร้อมทั้งคำนวณค่า AIC ของโครโมโซมที่ได้ ดังแสดงในรูป 3.4

ขั้นที่ 5 ทำการแปลงรหัสจากโครโมโซมที่ได้มาเป็นตัวแปรที่เหมาะสมจะอยู่ในตัวแบบถดถอย พร้อมทั้งคำนวณค่า AIC ทำการเลือกโครโมโซมที่ให้ค่า AIC น้อยที่สุด เพื่อนำมาเป็นตัวแบบที่เหมาะสมที่สุดสำหรับ ประชากรรุ่นที่ 1 ดังแสดงในรูป 3.5 โดยนำโครโมโซมพ่อแม่มาพิจารณาด้วย ดำเนินการทั้งสิ้นจำนวน 2,000 รุ่น แล้วนำโครโมโซมในรุ่นที่ 2,000 มาดำเนินการในขั้นต่อไป

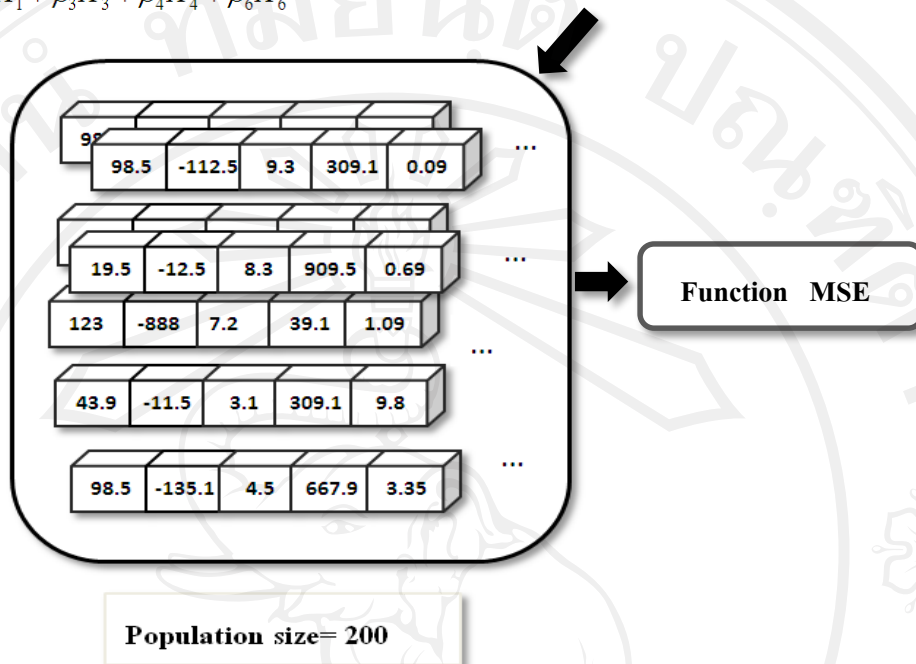
ขั้นที่ 6 เมื่อได้โครโมโซมที่ให้ค่า AIC ต่ำที่สุดแล้ว ทำการสร้างขอบเขตของการหาค่าสัมประสิทธิ์การถดถอยเพื่อทำการสุ่มหาค่าในการสร้างโครโมโซมในขั้นตอนต่อไป ซึ่งสร้างจากการประมาณค่าสัมประสิทธิ์การถดถอยของ OLS และค่า SE โดย ขอบล่างเท่ากับค่าสัมประสิทธิ์การถดถอยของ OLS - (2 x SE) ขอบบนเท่ากับค่าสัมประสิทธิ์การถดถอยของ OLS + (2 x SE)

ส่วนที่ 2

ขั้นที่ 1 จากโครโมโซมที่ได้ในส่วนที่ 1 ขั้นที่ 6 นำมาสร้างโครโมโซมแบบค่าจริงจากการสุ่มในช่วงขอบเขตที่กำหนดไว้ เพื่อเป็นประชากรเริ่มต้น โดยจำนวนยีนจะเท่ากับจำนวนตัวแปรอิสระที่ถูกเลือกให้อยู่ในตัวแบบที่เหมาะสมและ β_0 พร้อมทั้งแปลงรหัสเพื่อคำนวณค่า MSE ของแต่ละโครโมโซม ดังแสดงในรูป 3.6



$$y = \beta_0 + \beta_1 X_1 + \beta_3 X_3 + \beta_4 X_4 + \beta_6 X_6$$



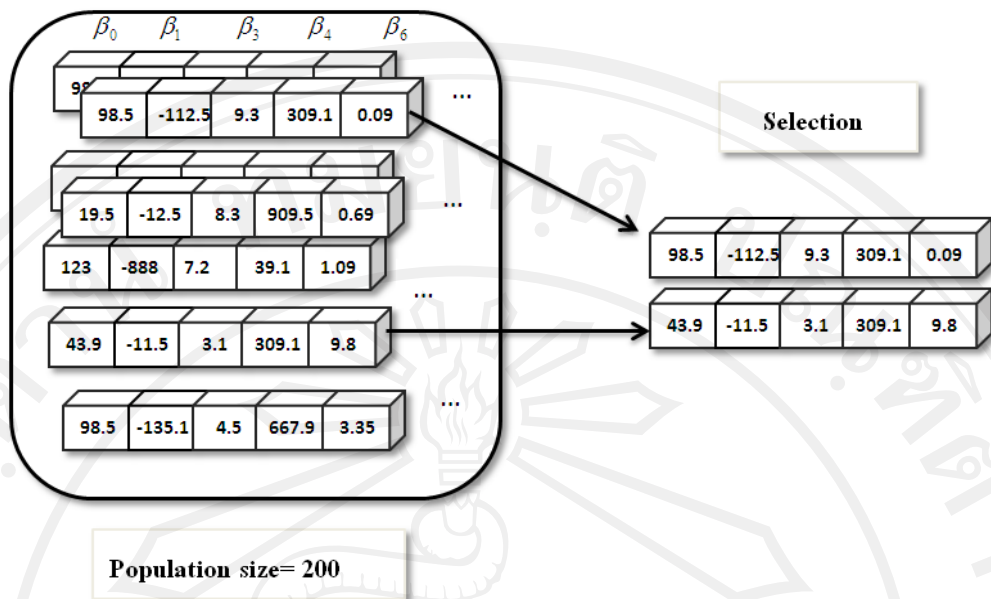
β_0	β_1	β_3	β_4	β_6
98.5	-112.5	9.3	309.1	0.09

$$\hat{y} = 98.5 - 112.5X_1 + 9.3X_3 + 309.1X_4 + 0.09X_6$$

$$MSE = 1599.88$$

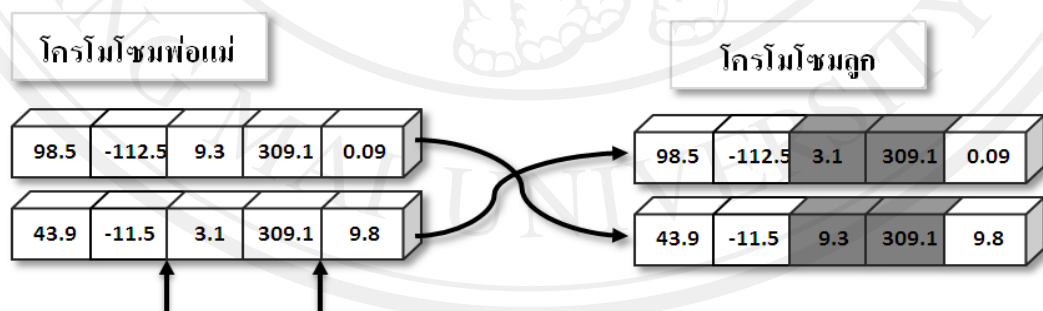
รูป 3.6 แสดงการสร้างประชากรเริ่มต้นในส่วนที่ 2 ของวิธี DGA และการแปลงรหัส

ขั้นที่ 2 ทำการคัดเลือกประชากรเพื่อเป็นโครโมโซมพ่อแม่ จากโครโมโซมที่มีค่า MSE ที่น้อยที่สุด 2 อันดับ ดังแสดงในรูป 3.7



รูป 3.7 แสดงการคัดเลือกโครโมโซมในส่วนที่ 2

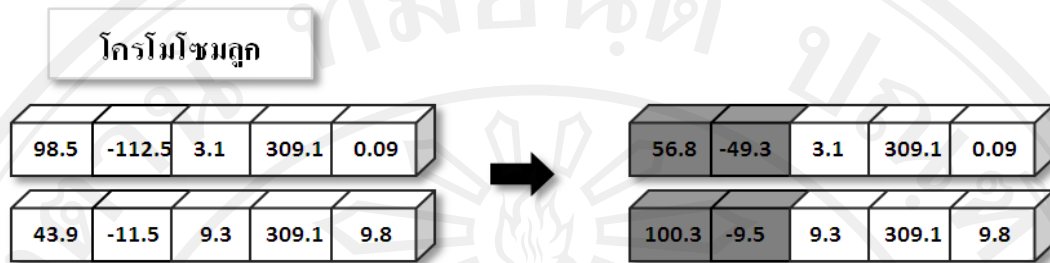
ขั้นที่ 3 กำหนดความน่าจะเป็นของการสลับสายพันธุ์ เท่ากับ 0.8 จากนั้นสุ่มตัวเลขในช่วง 0 ถึง 1 หากมีค่าน้อยกว่าหรือเท่ากับ 0.8 จะทำการสลับสายพันธุ์ ในที่นี้จะทำการสลับสายพันธุ์ 2 จุด จากนั้นสุ่มเพื่อเลือกตำแหน่งที่จะสลับสายพันธุ์ พร้อมทั้งทำการสลับสายพันธุ์ ดังแสดงในรูป 3.8



รูป 3.8 แสดงการสลับสายพันธุ์โครโมโซมในส่วนที่ 2

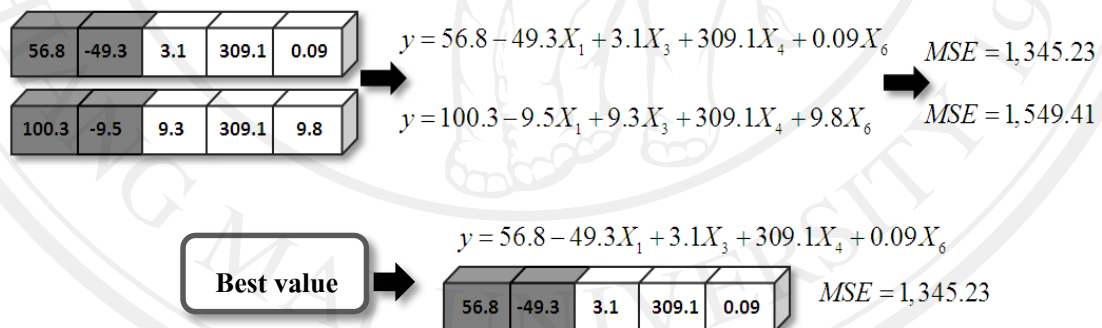
ขั้นที่ 4 กำหนดความน่าจะเป็นของการกลายพันธุ์ เท่ากับ 0.2 จากนั้นสุ่มตัวเลขในช่วง 0 ถึง 1 หากมีค่าน้อยกว่าหรือเท่ากับ 0.2 จะทำการกลายพันธุ์ ในกรณีที่มีการกลายพันธุ์ ตำแหน่งและจำนวนตำแหน่งที่จะทำการกลายพันธุ์เป็นไปโดย

ส่วน จากนั้นจึงทำการกลายพันธุ์ พร้อมทั้งคำนวณค่า MSE ของโครโมโซมที่ได้ ดังแสดงในรูป 3.9



รูป 3.9 แสดงการกลายพันธุ์โครโมโซมในส่วนที่ 2

ขั้นที่ 5 ทำการแปลงรหัสจากโครโมโซมที่ได้มาเป็นตัวแปรที่เหมาะสมจะอยู่ในตัวแบบถดถอยพร้อมทั้งคำนวณค่า MSE ทำการเลือกโครโมโซมที่ให้ค่า MSE น้อยที่สุด เพื่อนำมาเป็นตัวแบบที่เหมาะสมที่สุดสำหรับประชากรรุ่นที่ 1 โดยนำโครโมโซมพ่อแม่มาพิจารณาด้วย ดังแสดงในรูป 3.10 ดำเนินการทั้งสิ้นจำนวน 2,000 รุ่น



รูป 3.10 แสดงการแปลงรหัสของโครโมโซมของประชากรรุ่นที่ 1 และเลือกโครโมโซมในส่วนที่ 2

ขั้นที่ 6 นำโครโมโซมในรุ่นที่ 2,000 มาเป็นคำตอบในแต่ละรอบของการทำซ้ำในแต่ละสถานการณ์ที่สร้างขึ้น

3.2.3 การทำซ้ำในการดำเนินการคัดเลือกตัวแบบ

การกระทำซ้ำในการดำเนินการคัดเลือกตัวแบบของแต่ละสถานการณ์ โดยกำหนดจำนวนรอบการกระทำซ้ำและจำนวนรุ่นของประชากรที่ใช้ในกระบวนการของวิธี DGA และวิธี GA คำนวณจากสมการของการกำหนดจำนวนรอบของการประมวลผล ดังต่อไปนี้

กำหนดให้ $R_0=100$, $\alpha=0.05$, $t_{\frac{\alpha}{2}, R_0-1} = t_{0.025, 99} = 1.984$

และกำหนด $\varepsilon=150$ เนื่องจากการกำหนดจำนวนกระทำซ้ำเป็นการพิจารณาชุดข้อมูลที่แตกต่างกันในแต่ละรอบ ดังนั้นจึงต้องการให้มีความคลาดเคลื่อนไม่มากเกินไป

จะได้ว่า

$$S_0 = \sqrt{\frac{\sum_{i=1}^{R_0} (x_i - \bar{x})^2}{R_0 - 1}} = \sqrt{\frac{34,480,614.45}{99}} = 590.16$$

และ

$$\begin{aligned} R &\geq \left(\frac{t_{\frac{\alpha}{2}, R_0-1} S_0}{\varepsilon} \right)^2 \\ &\geq \left(\frac{(1.984)(590.16)}{150} \right)^2 \\ &\geq 61 \end{aligned}$$

พิจารณาจากผลงานวิจัยของนันทพร บุญสุข (7) ได้ทำการศึกษาเกณฑ์และสถิติทดสอบในการคัดเลือกตัวแปรในตัวแบบการถดถอยเชิงเส้นพหุคูณและจากค่า R ที่ได้จากการคำนวณ ดังนั้นจึงทำการกำหนดจำนวนกระทำซ้ำ 100 รอบ

การกำหนดจำนวนรุ่นของประชากรของวิธี DGA และวิธี GA มีดังนี้

กำหนดให้ $R_0=100$, $\alpha=0.05$, $t_{\frac{\alpha}{2}, R_0-1} = t_{0.025, 99} = 1.984$

และกำหนด $\varepsilon=5$ เนื่องจากการกำหนดจำนวนรุ่นของประชากรเป็นการพิจารณาในชุดข้อมูลเดียวกัน นั่นคือ ใช้ข้อมูลของ $X_{i1}, X_{i2}, \dots, X_{ik}$ และ y_i ในแต่ละรุ่นเหมือนกัน ดังนั้นจึงต้องการให้ค่าของฟังก์ชันเป้าหมายมีความใกล้เคียงกันมากที่สุดหรือแตกต่างกันน้อยที่สุด

จะได้ว่า

$$S_0 = \sqrt{\frac{\sum_{i=1}^{R_0} (x_i - \bar{x})^2}{R_0 - 1}} = \sqrt{\frac{714,770.19}{99}} = 84.97$$

และ

$$\begin{aligned} R &\geq \left(\frac{t_{\frac{\alpha}{2}, R_0 - 1} S_0}{\varepsilon} \right)^2 \\ &\geq \left(\frac{(1.984)(84.97)}{5} \right)^2 \\ &\geq 1,137 \end{aligned}$$

พิจารณาค่า R ที่ได้จากการคำนวณและการทดลองดำเนินการในจำนวนรุ่นที่
หลากหลาย ทำให้ทราบว่า การกำหนดจำนวนรุ่นประชากรของวิธี DGA และวิธี GA
ในจำนวน 2,000 รุ่น เหมาะสมที่สุด ดังนั้นจึงทำการกำหนดที่จำนวน 2,000 รุ่น