บทที่ 3
วิธีการดำเนินการวิจัย

งานวิจัยนี้มีจุดประสงค์เพื่อเปรียบเทียบวิธีการหาค่าคงที่ k วิธีต่างๆ ในการถดถอยแบบริดจ์ โดยข้อมูลที่ใช้ในการวิจัยครั้งนี้ได้มาจากการจำลองด้วยเทคนิคมอนติคาโลโดยใช้โปรแกรมภาษาฟอร์แทรน 77 ซึ่งกระทำซ้ำ 1,000 ครั้ง

3.1 แผนการวิจัย

กำหนดสถานการณ์ต่างๆ ที่ต้องการศึกษาดังต่อไปนี้

3.1.1 ขนาดตัวอย่างเท่ากับ 10 , 30 , 50 และ 100

3.1.2 จำนวนตัวแปรอิสระที่ใช้ เท่ากับ 3 , 5

3.1.3 ความคลาดเคลื่อนเป็นเลขสุ่มที่มาจากการแจกแจงแบบปกติ ซึ่งมีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ 0.05 , 0.5 , 1 และ 3

3.1.4 ระดับความสัมพันธ์ระหว่างตัวแปรอิสระ ตามกรณีต่างๆ ดังนี้

1) <u>กรณีจำนวนตัวแปรอิสระเท่ากับ 3</u>

พหุสัมพันธ์ระหว่างตัวแปรอิสระ 1 กลุ่ม ผู้วิจัยทำการศึกษาที่ระดับความสัมพันธ์ของกลุ่มตัวแปรมีค่าเท่ากับ (0.1) , (0.3) , (0.5) , (0.7) , (0.9) และ (0.99) [1] โดยแบ่งออกเป็น

ก. พหุสัมพันธ์ระหว่างตัวแปรอิสระ 2 ตัวแปร คือ X_1 และ X_2 มีความสัมพันธ์กัน

ข. พหุสัมพันธ์ระหว่างตัวแปรอิสระ 3 ตัวแปร คือ X_1, X_2 และ X_3 มีความสัมพันธ์กัน

2) <u>กรณีจำนวนตัวแปรอิสระเท่ากับ 5</u>

2.1) พหุสัมพันธ์ระหว่างตัวแปรอิสระ 1 กลุ่ม ผู้วิจัยทำการศึกษาที่ระดับความสัมพันธ์ของกลุ่มตัวแปรมีค่าเท่ากับ (0.1) , (0.3) , (0.5) , (0.7) , (0.9) และ (0.99) [1] โดยแบ่งออกเป็น

ก. พหุสัมพันธ์ระหว่างตัวแปรอิสระ 3 ตัวแปร คือ X_1, X_2 และ X_3 มีสัมพันธ์กัน

---

[1] ระดับความสัมพันธ์ของตัวแปรอิสระ 1 กลุ่ม ในกรณีต่าง ๆ

ข. พหุสัมพันธ์ระหว่างตัวแปรอิสระ 4 ตัวแปร คือ X_1, X_2, X_3 และ X_4 มีความสัมพันธ์กัน

ค. พหุสัมพันธ์ระหว่างตัวแปรอิสระ 5 ตัวแปร คือ X_1, X_2, X_3, X_4 และ X_5 มีความสัมพันธ์กัน

2.2) พหุสัมพันธ์ระหว่างตัวแปรอิสระ 2 กลุ่ม ผู้วิจัยทำการศึกษาที่ระดับความสัมพันธ์ของกลุ่มตัวแปรมีค่าเท่ากับ

(0.1,0.1) , (0.1,0.3) , (0.1,0.5) , (0.1,0.7) , (0.1,0.9) , (0.1,0.99) ,
(0.3,0.1) , (0.3,0.3) , (0.3,0.5) , (0.3,0.7) , (0.3,0.9) , (0.3,0.99) ,
(0.5,0.1) , (0.5,0.3) , (0.5,0.5) , (0.5,0.7) , (0.5,0.9) , (0.5,0.99) ,
(0.7,0.1) , (0.7,0.3) , (0.7,0.5) , (0.7,0.7) , (0.7,0.9) , (0.7,0.99) ,
(0.9,0.1) , (0.9,0.3) , (0.9,0.5) , (0.9,0.7) , (0.9,0.9) , (0.9,0.99) , และ
(0.99,0.1) , (0.99,0.3) , (0.99,0.5) , (0.99,0.7) , (0.99,0.9) , (0.99,0.99) [2]

โดยแบ่งออกเป็น

ก. พหุสัมพันธ์ระหว่างตัวแปรอิสระ 2 ตัวแปร 2 กลุ่ม ได้แก่ ตัวแปร X_1, X_2 มีพหุสัมพันธ์กัน และตัวแปร X_3 , X_4 มีพหุสัมพันธ์กัน

ข. พหุสัมพันธ์ระหว่างตัวแปรอิสระ 3 ตัวแปร 1 กลุ่ม และตัวแปรอิสระ 2 ตัวแปร 1 กลุ่ม ได้แก่ ตัวแปร X_1, X_2, X_3 มีพหุสัมพันธ์กัน และตัวแปร X_3 , X_4 มีพหุสัมพันธ์กัน

3.2 ขั้นตอนการวิจัย

3.2.1 สร้างข้อมูลของความคลาดเคลื่อนให้มีลักษณะตามที่ต้องการศึกษา

3.2.2 สร้างข้อมูลของตัวแปรอิสระ (X) ให้มีระดับความสัมพันธ์ตามที่กำหนด พร้อมทั้งค่าเฉพาะและเวกเตอร์เฉพาะของเมทริกซ์ $X'X$ และกำหนดให้ $X^* = XP$ และ $\beta^* = P'\beta$ โดยที่ P คือ เมทริกซ์ของเวกเตอร์เฉพาะ

3.2.3 สร้างข้อมูลของตัวแปรตาม (y) จากรูปแบบความสัมพันธ์ $y = X^*\beta^* + \varepsilon$

3.2.4 ประมาณค่าสัมประสิทธิ์การถดถอยจากวิธีการถดถอยแบบริดจ์ โดยการใช้วิธีการ

---

[2] ตัวเลขในวงเล็บแสดงถึงระดับความสัมพันธ์ของตัวแปรอิสระ 2 กลุ่มตัวแปร โดยที่ตัวแรกหมายถึง ระดับความสัมพันธ์ของตัวแปรอิสระในกลุ่มที่หนึ่ง และตัวที่สองหมายถึง ระดับความสัมพันธ์ของตัวแปรในกลุ่มที่สอง

ประมาณค่าคงที่ k ทั้ง 3 วิธี คือ วิธีโฮเอิร์นและเคนนาร์ด (HK) , วิธีค้นหาข้อมูลแบบลำดับ (SEQ), และวิธีเบส์(BAY) รวมทั้งหาค่าเฉลี่ยความคลาดเคลื่อนยกกำลังสองของตัวประมาณสัมประสิทธิ์การถดถอยที่ได้จากวิธีการต่างๆ

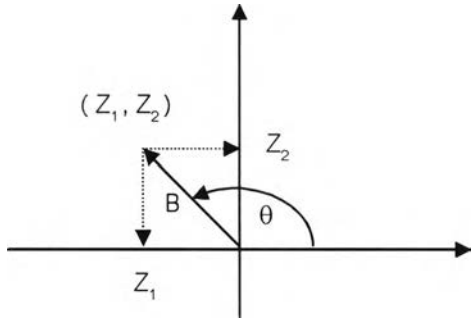
3.2.5 คำนวณค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนยกกำลังสอง(AMSE) ของตัวประมาณสัมประสิทธิ์การถดถอยที่ได้จากวิธีการต่างๆ และคำนวณเปอร์เซนต์ของอัตราส่วนผลต่างค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนยกกำลังสอง (PDMSE)พร้อมทั้งเปรียบเทียบและสรุปผลที่ได้

<u>รายละเอียดของขั้นตอนการวิจัย</u>

1) <u>การสร้างข้อมูลให้มีลักษณะการแจกแจงแบบปกติ</u> [3]

ในการสร้างข้อมูลให้มีลักษณะการแจกแจงแบบปกตินั้นจะใช้เลขสุ่ม (random number) ซึ่งมีการแจกแจงแบบสม่ำเสมอ (uniform) ในช่วง (0,1) เป็นพื้นฐาน ซึ่งรายละเอียดในการสร้างเลขสุ่ม มีดังนี้

การผลิตเลขสุ่มที่มีการแจกแจงแบบปกติจะใช้วิธีของ บ๊อกส์ มูลเลอร์ (Box Muller , 1958) โดยจะผลิตเลขสุ่มที่มีการแจกแจงแบบปกติมาตรฐาน ที่มีค่าเฉลี่ยเป็น 0 และความแปรปรวนเป็น 1 (N(0,1)) พร้อมกัน 2 ค่า และแต่ละค่าจะเป็นอิสระซึ่งกันและกัน โดยใช้ตัวผลิต (Generator) Z_1 และ Z_2 ดังรูปต่อไปนี้



จากรูปเราจะได้ว่า

$$Z_1 = B\cos\theta \quad \text{.....(3.1)}$$

$$Z_2 = B\sin\theta \quad \text{.....(3.2)}$$

---

[3] สินีนาถ กีอำไพ. การเปรียบเทียบอำนาจการทดสอบ สำหรับการทดสอบค่าสัมประสิทธ์สหสัมพันธ์ ในการแจกแจงปกติทวิและการแจกแจงปกติแกมมาทวิ. วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาสถิติ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย,2538.

เนื่องจาก $B^2 = Z_1^2 + Z_2^2$ มีการแจกแจงแบบไค-กำลังสอง (Chi - square distribution) ที่มีระดับขั้นความเสรีเท่ากับ 2 ซึ่งจะมีรูปแบบเหมือนกับการแจกแจงแบบชี้กำลัง (exponential distribution) ซึ่งมีค่าเฉลี่ยเท่ากับ 2 โดยการใช้วิธีการแปลงผกผัน (inverse - transformation) ได้ดังต่อไปนี้

$$B = \{-2 \ln (U)\}^{\frac{1}{2}} \qquad \text{.....(3.3)}$$

เมื่อ U เป็น เลขสุ่มที่มีการแจกแจงแบบสม่ำเสมอ

จากการสมมาตรของการแจกแจงปกติ จะได้ว่า มีการแจกแจงสม่ำเสมอระหว่าง 0 ถึง 2π เรเดียน และรัศมี B กับมุม θ เป็นอิสระต่อกัน จากสมการที่ (3.1) , (3.2) และ (3.3) เราสามารถสร้างเลขสุ่มที่มีการแจกแจงแบบปกติมาตรฐานจากเลขสุ่ม 2 ชุด คือ U_1 และ U_2 กล่าวคือ

$$Z_1 = \{-2\ln(U_1)\}^{\frac{1}{2}} \cos(2\pi U_2)$$
$$Z_2 = \{-2\ln(U_1)\}^{\frac{1}{2}} \sin(2\pi U_2)$$

ซึ่ง Z_1 และ Z_2 เป็นเลขสุ่มที่สร้างจาก SUBROUTINE RANDOM เมื่อได้เลขสุ่มที่มีการแจกแจงแบบปกติมาตรฐานแล้ว จะทำการแปลงค่าเลขสุ่มดังกล่าวโดยที่

$$NZ = \mu + \sigma Z_1$$
$$NZ = \mu + \sigma Z_2$$

จะได้ว่า NZ มีการแจกแจงแบบปกติ ด้วยค่าเฉลี่ยเท่ากับ μ และความแปรปรวน เป็น σ^2

2. <u>การสร้างข้อมูลให้มีความสัมพันธ์เชิงเส้นตรง</u>

ในการวิจัยครั้งนี้จะสร้างข้อมูลของตัวแปรอิสระ $\underset{\sim}{X} = (X_1, X_2, \dots, X_p)'$ ให้มีการแจกแจงปกติหลายตัวแปร (multivariate normal distribution) ที่มีค่าพารามิเตอร์ $\underset{\sim}{\mu} = 0$, Σ เป็นเมทริกซ์ความแปรปรวนร่วม(covariance matrix) ที่แสดงความสัมพันธ์ระหว่างตัวแปรอิสระ X สามารถเขียนได้เป็น $\underset{\sim}{X}_{pxl} \sim N_p(\underset{\sim}{\mu}, \Sigma)$ โดยที่

$$\Sigma = E\left[\left(\underset{\sim}{X}-\underset{\sim}{\mu}\right)\left(\underset{\sim}{X}-\underset{\sim}{\mu}\right)'\right] = \begin{bmatrix} \sigma_{11} & \cdots & \cdots & \sigma_{1p} \\ \vdots & \sigma_{22} & & \vdots \\ \vdots & & \ddots & \vdots \\ \sigma_{p1} & \cdots & \cdots & \sigma_{pp} \end{bmatrix}$$

ในขั้นเริ่มต้นเราจะคำนวณหาเมทริกซ์ lower Triangular (C_{pxp}) ที่ทำให้ $\Sigma = CC'$ จากนั้นจะ

1) สร้างเวกเตอร์ Z ที่ประกอบด้วยตัวแปรที่มีการแจกแจงแบบปกติมาตรฐานที่เป็นอิสระซึ่งกันและกัน p ตัว

2) คำนวณ $\underset{\sim}{X} = C\underset{\sim}{Z} + \underset{\sim}{\mu}$ เนื่องจาก Σ เป็นเมทริกซ์สมมาตร จะใช้วิธีรากที่สอง (square root method) ในการคำนวณเมทริกซ์ C ดังนี้

จาก $\underset{\sim}{X} = C\underset{\sim}{Z} + \underset{\sim}{\mu}$

จะได้ว่า $X_1 = C_{11} Z_1 + \mu_1$

$$Var(X_1) = \sigma_{11} = C_{11}^2$$

$$X_2 = C_{21} Z_1 + C_{22} Z_2 + \mu_2$$

$$\therefore Var(X_2) = \sigma_{22} = Var(C_{21} Z_1 + C_{22} Z_2)$$

$$\because E\left[(X_1 - \mu_1)(X_2 - \mu_2)\right] = \sigma_{12} = E(C_{11}Z_1 + C_{21}Z_1 + C_{22}Z_2)$$

ดังนั้น

$$C_{21} = \frac{\sigma_{12}}{C_{11}} = \frac{\sigma_{12}}{\sqrt{\sigma_{11}}}$$

$$C_{22} = \sqrt{\sigma_{22} - \frac{\sigma_{21}^2}{\sigma_{11}}}$$

รูปแบบทั่วไปของ C_{ij} แสดงได้ดังนี้คือ

$$C_{ij} = \frac{\sigma_{ij} - \sum_{k=1}^{j-1} C_{ik}C_{kj}}{\sqrt{\sigma_{jj} - \sum_{k=1}^{j-1} C_{jk}}}$$

เมื่อ $\sum_{k=1}^{0} C_{ik}C_{jk} = 0 \quad ; \quad 1 \le j \le i \le p$

เราสามารถพิสูจน์ได้ว่า

$$\begin{aligned}\Sigma &= E\left[(CZ)(CZ)'\right] \\ &= E\left[CZZ'C'\right] \\ &= CE(ZZ')C' = CC'\end{aligned}$$

3. <u>การสร้างข้อมูลของตัวแปรตาม (y)</u>

ก่อนที่จะทำการสร้างข้อมูลของตัวแปรตามนั้น เราจะทำการหาค่าเฉพาะและเวกเตอร์เฉพาะของเมทริกซ์ $X'X$

$$\because \quad X^* = XP$$

และ $\underset{\sim}{\beta}^* = P'\underset{\sim}{\beta}$

$$\therefore \quad X^{*\prime}X^* = P'X'XP = \Lambda = diag(\lambda_i)$$

เมื่อ λ เป็น ค่าเฉพาะ

และ P เป็น เมทริกซ์ของเวกเตอร์เฉพาะ ซึ่งมีคุณสมบัติเป็นเมทริกซ์เชิงตั้งฉาก

จากนั้นจะสร้างข้อมูลของตัวแปรตาม จากตัวแบบ $\underset{\sim}{y} = X^{*}\underset{\sim}{\beta}^{*} + \underset{\sim}{\varepsilon}$ โดยจะกำหนดค่าสัมประสิทธิ์การถดถอยพหุคูณ($\underset{\sim}{\beta}^{*}$) จากเวกเตอร์เฉพาะที่สอดคล้องกับค่าเฉพาะที่มีค่ามากที่สุดและน้อยที่สุด

4. <u>การหาค่าเฉลี่ยความคลาดเคลื่อนยกกำลังสอง</u>

4.1 การคำนวณค่าเฉลี่ยความคลาดเคลื่อนยกกำลังสองของตัวประมาณสัมประสิทธิ์การถดถอยจากวิธีการถดถอยริดจ์ ($MSE(\underset{\sim}{\hat{\beta}}^{*}_{R})$) โดยใช้การหาค่า k จากวิธีต่างๆ ดังที่กล่าวไว้แล้วในบทที่ 2 จะได้ว่า

$$MSE(\underset{\sim}{\hat{\beta}}^{*}_{R}) = \frac{\sum_{i=1}^{p}(\underset{\sim}{\hat{\beta}}^{*}_{R} - \underset{\sim}{\beta}^{*})^{2}}{p}$$

เมื่อ $\underset{\sim}{\hat{\beta}}^{*}_{R}$ คือ ตัวประมาณสัมประสิทธิ์การถดถอยพหุคูณจากวิธีการถดถอย ริดจ์ โดยใช้การหาค่า k จากวิธีต่างๆ

และ $\underset{\sim}{\beta}^{*}$ คือ สัมประสิทธิ์การถดถอยพหุคูณ

4.2 การคำนวณค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนยกกำลังสอง(AMSE) ของตัวประมาณสัมประสิทธิ์การถดถอยพหุคูณจากวิธีการถดถอย ริดจ์ เนื่องจากการวิจัยครั้งนี้ได้ทำการจำลองข้อมูลด้วยเทคนิคมอนติคาโลกระทำซ้ำ 1000 ครั้ง ดังนั้น ค่า AMSE ของแต่ละวิธี คือ

$$AMSE = \frac{1}{1000}\sum_{j=1}^{1000} MSE(\underset{\sim}{\hat{\beta}}^{*}_{R(j)})$$

4.3 การคำนวณค่าเปอร์เซนต์ของอัตราส่วนผลต่างค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนยกกำลังสอง(PDMSE)

$$PDMSE = \frac{AMSE_{(i)} - AMSE_{(\min)}}{AMSE_{(\min)}}$$

เมื่อ $AMSE_{(\min)}$ คือ ค่า AMSE ของวิธีที่ให้ค่าน้อยที่สุด

และ $AMSE_{(i)}$ คือ ค่า AMSE ของวิธีที่ i ; $i = 1, 2, 3$

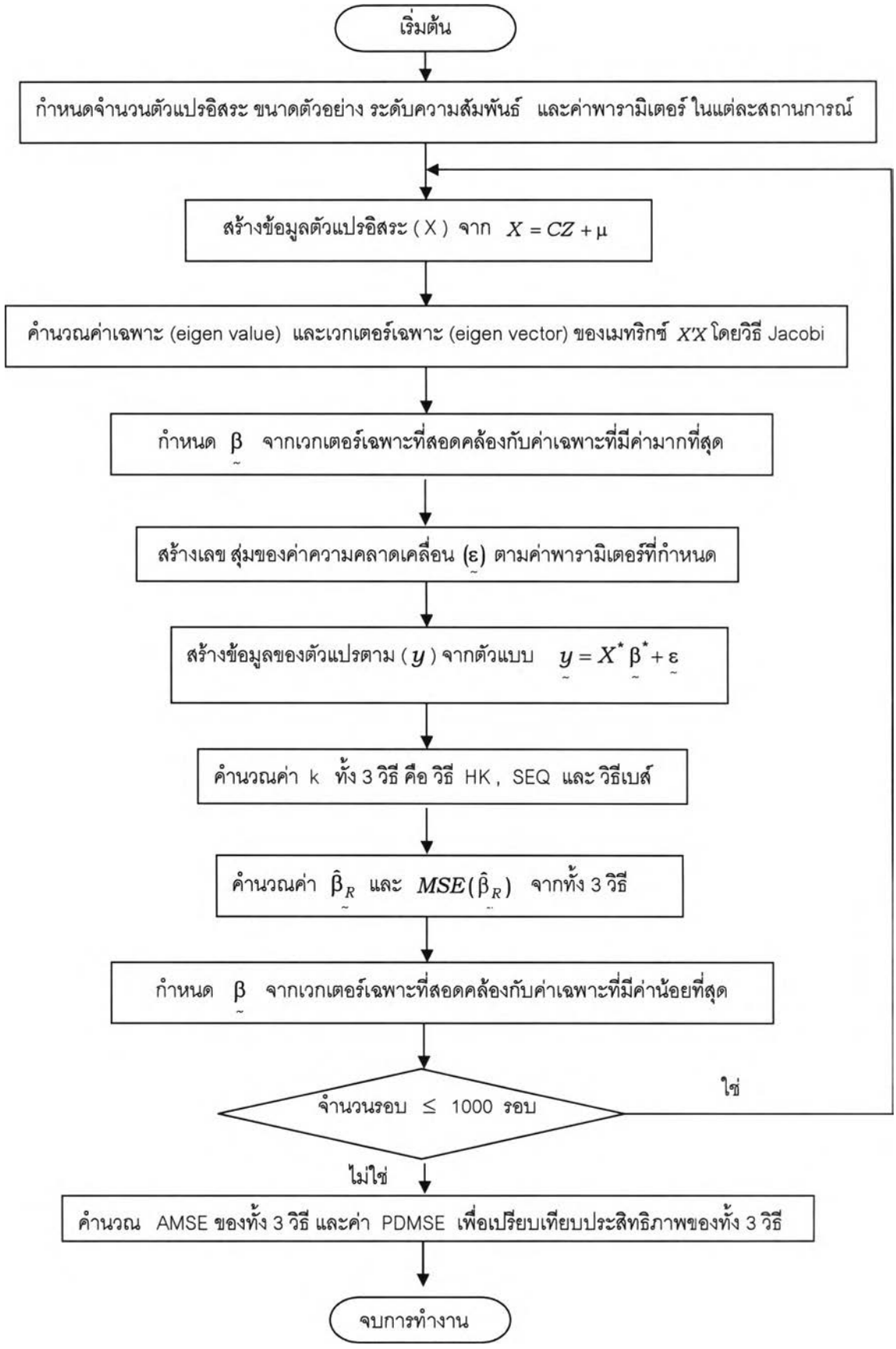
ตารางที่ 3.1 ลักษณะการทำงานของโปรแกรมทั้งหมดที่ใช้ในการวิจัย

อันดับที่	ชื่อโปรแกรม	คุณสมบัติของโปรแกรม	ชื่อโปรแกรมและฟังก์ชันที่เรียกใช้
โปรแกรมหลัก			
1	MAIN	คำนวณค่า MSE , AMSE ของตัวประมาณสัมประสิทธิ์การถดถอยแบบริดจ์และคำนวณ ค่า PDMSE	MATRIX_C , GENX ERROR , JACOBI , VECT,GENY,OLS, HK,BAY, SEQUEN
โปรแกรมย่อย			
1	MATRIX_C	คำนวณเมทริกซ์ C เพื่อใช้ในการสร้างข้อมูลตัวแปรอิสระให้มีความสัมพันธ์กัน	
2	GENX	สร้างข้อมูลตัวแปรอิสระ และคำนวณเมทริกซ์ $X'X$	MULTI , NORMAL
3	RANDOM	สร้างเลขสุ่มที่มีการแจกแจงสม่ำเสมอ	
4	MULTI	สร้างข้อมูลตัวแปรอิสระให้มีความสัมพันธ์กัน	
5	JACOBI	คำนวณค่าเฉพาะ และ เวกเตอร์เฉพาะโดยวิธี Jacobi	
6	GENY	สร้างข้อมูลของตัวแปรตาม	
7	INVRS	คำนวณเมทริกซ์ผกผัน	
8	OLS	คำนวณค่าสัมประสิทธิ์การถดถอยพหุคูณและค่าความแปรปรวนจากวิธีกำลังสองน้อยสุด	INVRS
9	ERROR	สร้างเลขสุ่มของค่าความคลาดเคลื่อน	NORMAL
10	HK	คำนวณค่า k , สัมประสิทธิ์การถดถอยพหุคูณและค่าเฉลี่ยความคลาดเคลื่อนยกกำลังสองจากวิธีการถดถอยริดจ์ แบบวิธี HK	SORT

ตารางที่ 3.1 (ต่อ) แสดงลักษณะการทำงานของโปรแกรมทั้งหมดที่ใช้ในการวิจัย

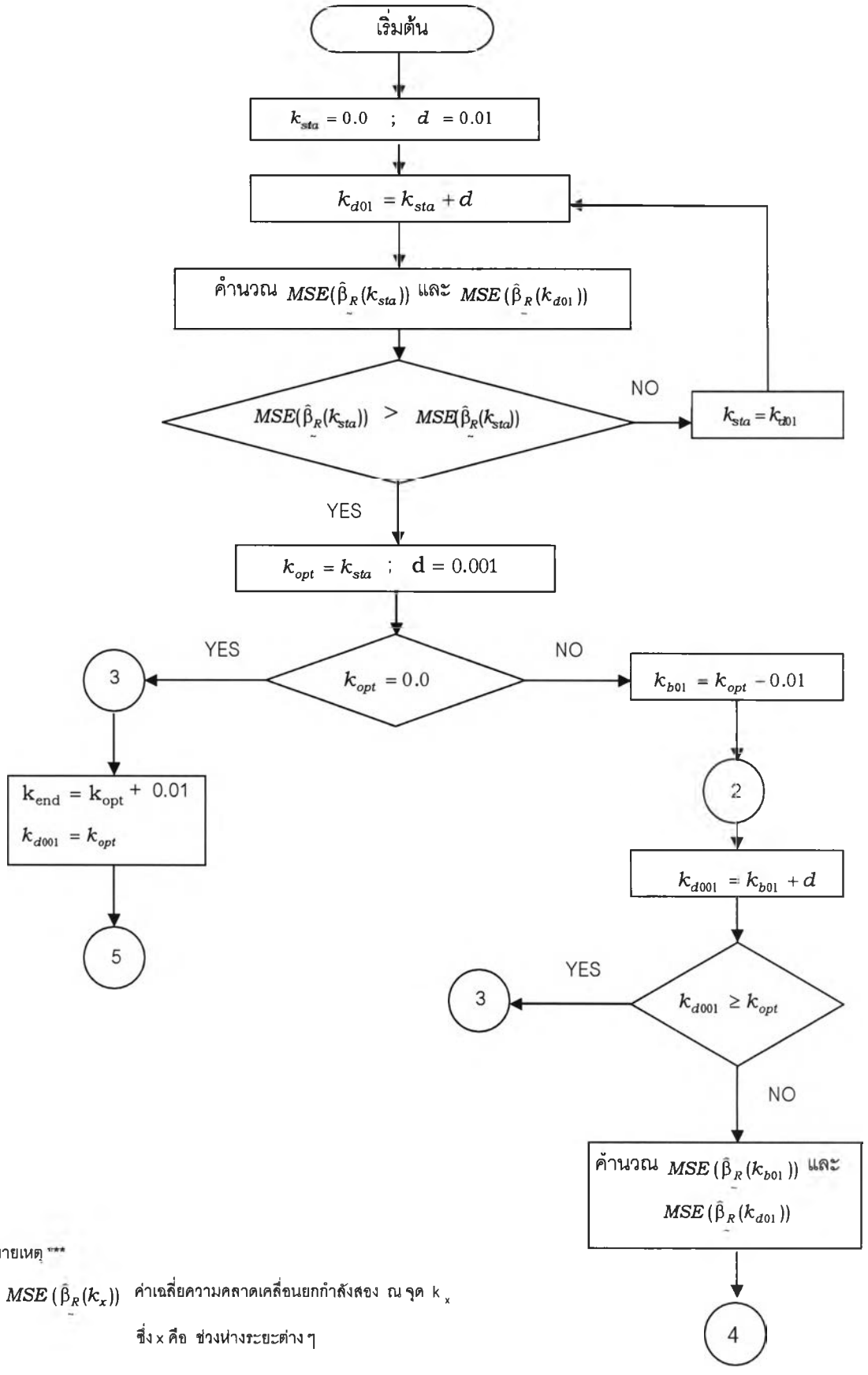
อันดับที่	ชื่อโปรแกรม	คุณสมบัติของโปรแกรม	ชื่อโปรแกรมและฟังก์ชันที่เรียกใช้
โปรแกรมย่อย			
11	BAY	คำนวณค่า k , สัมประสิทธิ์การถดถอยพหุคูณและค่าเฉลี่ยความคลาดเคลื่อนยกกำลังสองจากวิธีการถดถอยริดจ์แบบวิธีเบส์	
12	SEQU	คำนวณค่า k โดยวิธีค้นหาข้อมูลแบบลำดับ	MSES
13	SORT	เรียงลำดับค่าสัมประสิทธิ์การถดถอยพหุคูณจากวิธีกำลังสองน้อยสุด	
14	VECT	คำนวณค่าพารามิเตอร์จากเวกเตอร์เฉพาะ	
ฟังก์ชัน			
1	NORMAL	สร้างเลขสุ่มที่มีการแจกแจงปกติ	RAND
2	MSES	คำนวณค่าเฉลี่ยของความคลาดเคลื่อนยกกำลังสอง	

แผนผังการเขียนโปรแกรม



I20975673

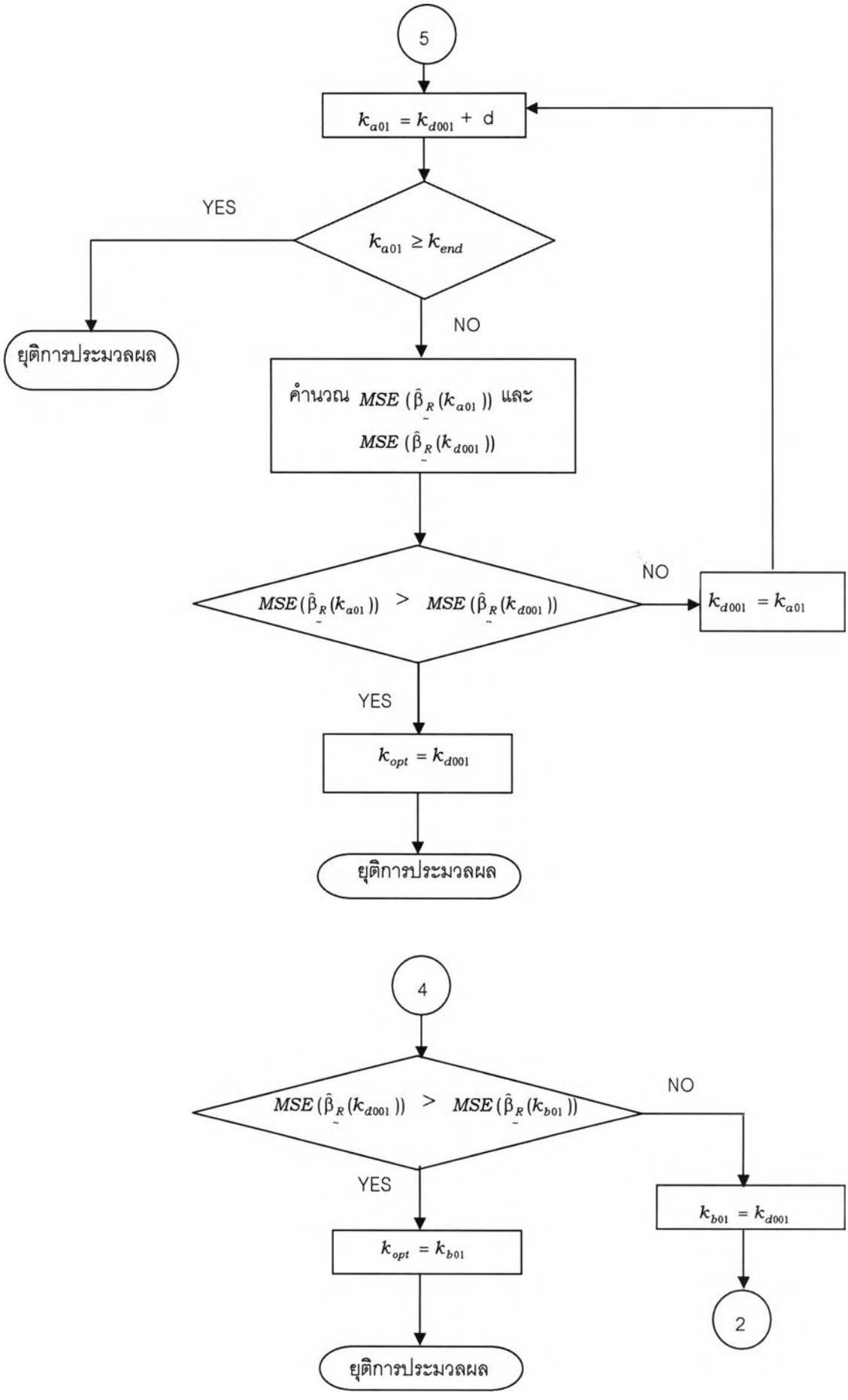
แผนผังการหาค่า k โดยวิธี ค้นหาข้อมูลแบบลำดับ



*** หมายเหตุ ***

$MSE(\hat{\beta}_R(k_x))$ ค่าเฉลี่ยความคลาดเคลื่อนยกกำลังสอง ณ จุด k_x

ซึ่ง x คือ ช่วงห่างระยะต่าง ๆ


5
$k_{a01} = k_{d001} + d$
YES
$k_{a01} \geq k_{end}$
NO
ยุติการประมวลผล
คำนวณ $MSE(\tilde{\beta}_R(k_{a01}))$ และ $MSE(\tilde{\beta}_R(k_{d001}))$
$MSE(\tilde{\beta}_R(k_{a01})) > MSE(\tilde{\beta}_R(k_{d001}))$
NO
$k_{d001} = k_{a01}$
YES
$k_{opt} = k_{d001}$
ยุติการประมวลผล
4
$MSE(\tilde{\beta}_R(k_{d001})) > MSE(\tilde{\beta}_R(k_{b01}))$
NO
YES
$k_{b01} = k_{d001}$
$k_{opt} = k_{b01}$
2
ยุติการประมวลผล